

# **The Role of GATA-1 Isoforms in Haematopoiesis**

**Dr Christina Halsey B.M., B.Ch.,  
B.A.(Hons).**



This thesis is submitted to the University of Glasgow  
for the degree of Doctor of Philosophy

Division of Immunology, Infection and Inflammation  
Faculty of Medicine  
University of Glasgow

February 2009

## Summary

---

GATA-1 is a key haematopoietic transcription factor which plays a pivotal role in differentiation of the erythroid, megakaryocytic, eosinophilic, mast cell and dendritic cell lineages. Since its initial cloning and characterisation in 1989 a huge amount of information has been gathered on the molecular mechanisms of action of GATA-1. This knowledge has helped understanding of the processes by which cells enact differentiation programmes and suppress alternative lineage choices. GATA-1 produces at least two protein isoforms – the well characterised GATA-1 full-length (GATA-1FL) isoform and a truncated isoform – GATA-1 short (GATA-1s). GATA-1FL comprises two conserved Zinc fingers (which interact with DNA and essential co-factors), a C-terminal tail (of mostly unknown function) and an N-terminal domain (thought to confer activation properties to the molecule, but which may also be involved in transcriptional repression). GATA-1s lacks the N-terminal domain but is otherwise identical. The biological role of GATA-1s is unknown and this isoform received scant attention until the discovery that GATA-1FL mutations were linked to a rare, but highly informative, acute megakaryoblastic leukaemia seen in children with Down syndrome (constitutional trisomy 21). This discovery was particularly interesting, not only because the association between trisomy 21 and the X-linked GATA-1 mutation was extremely tight (being seen in 100% of the cases examined), but also because the GATA-1FL mutations were not randomly located, but rather clustered within the N-terminus, allowing unhindered production of the GATA-1s isoform. This finding led to interest in the pathological and physiological role of GATA-1s in haematopoiesis.

Some insight has been gained into the pathological role of GATA-1s by creation of a GATA-1s knock-in transgenic mouse and by experiments looking at the ability of GATA-1s to rescue GATA-1 deficient embryonic stem (ES) cell lines. GATA-1s produces hyperproliferation of fetal liver meg-erythroid progenitors but allows at least partial differentiation of these cells. However, a number of key questions remain. In particular what is the physiological role of GATA-1s and the reason for the tight association between trisomy 21 and GATA-1s mutations?

Given this background, this thesis describes experiments designed to address the physiological role of GATA-1s, to establish whether additional GATA-1 isoforms exist, and to investigate the association between GATA-1 isoform expression and trisomy 21.

Firstly a comprehensive expression analysis was performed in murine and human primary tissues and cell lines. This aimed to identify whether GATA-1s had a unique expression profile, either in particular lineages, or at distinct stages of haematological ontogeny. Reverse-transcriptase polymerase chain reaction (RT-PCR) and western blot analyses showed that the expression patterns of GATA-1s and GATA-1FL were virtually identical, with the possible exception of one human primary monocytic cell preparation which appeared to preferentially express GATA-1s.

Before proceeding to further analysis of GATA-1s a search was made for additional GATA-1 isoforms using *in silico* analysis, RT-PCR and western blotting. This led to identification of a clone carrying a GATA-1 mutation involving the C-terminal tail, derived from a patient with chronic myeloid leukaemia. An analysis of the properties of this clone was performed, confirming its altered C-terminus and demonstrating that this conferred increased transactivation properties on the molecule as measured by luciferase assays. This observation suggests that the C-terminal tail may be an important, and previously under-recognised, functional region of the GATA-1 molecule. The discovery of this potentially hyper-functioning GATA-1 mutation led to investigation of whether GATA-1 mutations could be a widespread phenomenon in CML. However, GATA-1 mutational analysis in 21 patient samples from CML blast crisis did not reveal any additional coding mutations.

To address the physiological role of GATA-1s, attempts were made to perform gene targeting in murine embryonic stem cells to produce isoform specific knock-out cells *i.e.* ES cells engineered so that they exclusively express the GATA-1FL isoform (a GATA-1s knock-out) or the GATA-1s isoform (a GATA-1FL knockout). These cells could then be used in *in vitro* haematopoietic differentiation assays and for transcriptional profiling. In this way it was hoped to establish whether GATA-1s fulfilled any unique roles in primitive or definitive haematopoiesis that could not be compensated for by the presence of the GATA-1FL isoform. Unfortunately, despite evidence of apparently successful targeting from PCR screening of ES cell clones, it was impossible to confirm the existence of endogenously targeted alleles on Southern blotting. Following exhaustive attempts at screening further clones and subclones (more than 1000 clones in total), this approach was abandoned in favour of transgenic expression of GATA-1 isoforms in cell lines.

Transgenic expression studies in murine ES cells showed that whilst GATA-1FL expression led to an expansion in numbers and maturity of erythroid and non-erythroid haematopoietic colonies *in vitro*, GATA-1s was incapable of supporting colony formation

in this assay. Studies then moved on to human cell lines. Two cell lines were identified, both capable of in vitro haematopoietic differentiation into megakaryocytic and erythroid cells, but one carrying trisomy 21 (Meg-01) and the other disomic for chromosome 21 (K562). GATA-1FL expression in these cells generally drove differentiation along the megakaryocytic or erythroid lineage as measured by DNA ploidy analysis, haemoglobinisation, upregulation of erythroid or megakaryocytic gene expression (by quantitative PCR) and suppression of alternative lineage genes (PU.1 and Ikaros) and genes associated with progenitor proliferation (cyclin D2 and c-myb). GATA-1s, in contrast, produced less evidence of differentiation with lower DNA ploidy, less up-regulation of erythroid genes and failure to repress other lineage and haematopoietic progenitor associated genes. Examination of the link with trisomy 21 confirmed that the chromosome 21 candidate gene *Erg3* was upregulated in trisomic cells and that expression of GATA-1s appeared to confer a selective advantage in the presence of trisomy 21. However, no clear mechanistic reasons for the selective advantage could be identified.

Overall, these studies show widespread GATA-1s expression in haematopoietic cells, confirm the association with inadequate repression of genes associated with primitive progenitors, and suggest that the C-terminal tail of GATA-1 may be an important functional part of the molecule. Finally, these observations have generated a number of testable hypotheses which could form the basis for future work.

# Contents

---

<b>SUMMARY .....</b>	<b>2</b>
<b>TABLE OF CONTENTS .....</b>	<b>5</b>
<b>LIST OF TABLES .....</b>	<b>12</b>
<b>LIST OF FIGURES .....</b>	<b>13</b>
<b>LIST OF PUBLICATIONS ARISING FROM THIS WORK.....</b>	<b>16</b>
<b>ACKNOWLEDGEMENT.....</b>	<b>17</b>
<b>DECLARATION .....</b>	<b>18</b>
<b>ABBREVIATIONS.....</b>	<b>19</b>
<b>1 INTRODUCTION.....</b>	<b>21</b>
<b>1.1 HAEMATOPOIESIS .....</b>	<b>21</b>
1.1.1 HAEMATOPOIESIS DURING DEVELOPMENT .....	23
1.1.1.1 Extra-embryonic and intra-embryonic haematopoiesis .....	23
1.1.1.2 Primitive and definitive haematopoiesis.....	26
1.1.1.3 Haematopoietic niches.....	28
1.1.2 THE HAEMATOPOIETIC HIERARCHY .....	30
1.1.2.1 The haematopoietic stem cell .....	30
1.1.2.2 Progenitor cells.....	30
1.1.3 LINEAGE SPECIFICATION AT THE MOLECULAR LEVEL .....	31
1.1.3.1 The role of cytokines in differentiation .....	32
1.1.3.2 The role of transcription factors in differentiation.....	33
1.1.3.2.1 Master regulators .....	34
1.1.3.2.2 Concentration dependent effects.....	35
1.1.3.2.3 Negative as well as positive effects on differentiation .....	35
1.1.3.2.4 The importance of cellular context and timing of expression .....	36
1.1.3.2.5 Altering specificity by use of isoforms.....	36
1.1.3.3 Other methods of control of gene expression .....	37
1.1.3.3.1 Epigenetic regulation of gene expression .....	37
1.1.3.3.2 Small non-coding RNAs.....	38
1.1.3.3.3 Post-transcriptional control of gene expression.....	39
<b>1.2 GATA-1: A KEY HAEMATOPOIETIC TRANSCRIPTION FACTOR .....</b>	<b>40</b>
1.2.1 THE GATA FAMILY OF TRANSCRIPTION FACTORS .....	40
1.2.2 CLONING AND INITIAL CHARACTERISATION OF GATA-1 .....	40
1.2.3 GATA-1 PLAYS A KEY ROLE IN ERYTHROPOIESIS .....	44
1.2.3.1 Effects of GATA-1 deficiency in vitro.....	44
1.2.3.2 GATA-1 knockout embryos die due to severe anaemia .....	45
1.2.3.3 Knockdown variants allow survival of embryos .....	45
1.2.4 MECHANISMS OF GATA-1 ACTIVITY .....	46
1.2.4.1 GATA-1 Cofactors .....	47
1.2.4.2 Regulators of GATA-1 expression .....	49
1.2.5 THE ROLE OF GATA-1 IN MEGAKARYOCYTE DEVELOPMENT .....	51
1.2.5.1 Generation of transgenic mice with selective loss of GATA-1 expression in the megakaryocytic lineage .....	51
1.2.5.2 Gene expression mirrors that seen in the erythroid lineage.....	52
1.2.5.3 Rescue of proliferation and differentiation defects can be uncoupled and mapped to different domains of GATA-1 .....	52
1.2.6 ROLE IN OTHER LINEAGES .....	53

1.2.6.1	Eosinophils .....	53
1.2.6.2	Mast Cells .....	53
1.2.6.3	Dendritic cells .....	54
1.2.6.4	Sertoli cells .....	54
1.2.7	GATA-1 ALSO PLAYS ROLES IN LINEAGE COMMITMENT .....	54
1.2.8	GATA-1 ISOFORMS .....	55
1.2.8.1	GATA-1s is produced by alternative translation in mice .....	55
1.2.8.2	GATA-1s shows reduced transactivation potential in vitro .....	56
1.2.8.3	Creation of a transgenic mouse expressing GATA-1s .....	56
1.2.8.4	Insights from a naturally occurring human mutation .....	58
1.2.8.5	Summary of known GATA-1s function .....	59
<b>1.3</b>	<b>LEUKAEMIA .....</b>	<b>59</b>
1.3.1	THE GENETIC BASIS FOR LEUKAEMIA DEVELOPMENT .....	60
1.3.1.1	Class 1 and Class 2 mutations .....	60
1.3.1.2	In utero acquisition of mutations .....	61
1.3.1.3	The cellular origins of leukaemia .....	63
1.3.1.4	Primary and secondary changes in leukaemic cells .....	64
1.3.2	DOWN SYNDROME AND LEUKAEMIA .....	65
1.3.2.1	TAM and AMKL - natural history and clinical features .....	65
1.3.2.2	GATA-1 mutations and DS TAM/AMKL .....	68
1.3.2.3	Evidence for a fetal liver progenitor cell population as the target for GATA-1 mutations 71	
1.3.2.4	First genetic hit – Trisomy 21 .....	72
1.3.2.5	Third genetic hit - mutations in Jak 2 and 3? .....	76
1.3.3	CHRONIC MYELOID LEUKAEMIA .....	78
1.3.3.1	Molecular pathology .....	78
1.3.3.2	Similarities with transient abnormal myelopoiesis .....	79
1.3.3.3	The role of transcription factors in progression of disease .....	79
<b>1.4</b>	<b>AIMS AND OBJECTIVES .....</b>	<b>80</b>
<b>2</b>	<b>MATERIALS AND METHODS .....</b>	<b>82</b>
<b>2.1</b>	<b>GENERAL SOLUTIONS AND CONSUMABLES .....</b>	<b>82</b>
<b>2.2</b>	<b>MEDIA .....</b>	<b>86</b>
<b>2.3</b>	<b>BACTERIAL CULTURE MEDIA .....</b>	<b>88</b>
<b>2.4</b>	<b>PRIMERS, RNAI REAGENTS AND ANTIBODIES .....</b>	<b>88</b>
2.4.1	GATA-1 PRIMERS .....	88
2.4.2	PRIMERS USED FOR VECTOR CONSTRUCTION .....	89
2.4.3	PRIMERS FOR SITE-DIRECTED MUTAGENESIS .....	89
2.4.4	PRIMERS FOR RT-PCR TRANSCRIPTIONAL PROFILING .....	89
2.4.5	PRIMERS FOR CML MUTATIONAL ANALYSIS .....	90
2.4.5.1	Direct sequencing: .....	90
2.4.5.2	WAVE .....	90
2.4.6	PRIMERS AND PROBES FOR ES CELL TARGETING .....	91
2.4.7	PROBE/PRIMER PAIRS FOR REAL-TIME PCR .....	91
2.4.8	RNAI DUPLEXES .....	91
2.4.9	ANTIBODIES .....	92
2.4.9.1	For Western blotting .....	92
2.4.9.2	Antibodies for flow cytometry .....	92
<b>2.5</b>	<b>CELL CULTURE METHODS .....</b>	<b>93</b>
2.5.1	THAWING OF CELL LINES FROM FROZEN STOCKS .....	93
2.5.2	CELL COUNTING .....	93
2.5.3	MAINTENANCE OF CELL LINES IN CULTURE .....	94
2.5.3.1	Passage of adherent cell cultures .....	94
2.5.3.2	Passage of suspension cell cultures .....	94
2.5.4	FREEZING DOWN OF CELL LINES .....	94
<b>2.6</b>	<b>ANIMAL METHODS .....</b>	<b>95</b>
2.6.1	ISOLATION OF MOUSE BONE MARROW .....	95

2.6.2	MOUSE EMBRYONIC DISSECTIONS .....	95
<b>2.7</b>	<b>MOLECULAR BIOLOGY METHODS: DNA .....</b>	<b>96</b>
2.7.1	PREPARATION OF GENOMIC DNA .....	96
2.7.2	PURIFICATION, CONCENTRATION AND QUANTIFICATION OF DNA .....	96
2.7.3	RESTRICTION ENZYME DIGESTS .....	97
2.7.4	GEL ELECTROPHORESIS OF DNA .....	98
2.7.5	GEL EXTRACTION AND PCR PURIFICATION .....	99
2.7.6	SEQUENCING OF DNA .....	99
2.7.7	SOUTHERN BLOTTING .....	100
2.7.7.1	DNA digestion and gel electrophoresis .....	100
2.7.7.2	Preparation of the gel and membrane for transfer .....	100
2.7.7.3	Assembly of transfer apparatus and transfer of DNA .....	101
2.7.7.4	Preparation of <sup>32</sup> P labelled DNA probe .....	102
2.7.7.5	Probing of membrane for complementary DNA sequence .....	102
2.7.7.6	Stripping the Hybond-XL membrane .....	103
<b>2.8</b>	<b>MOLECULAR BIOLOGY METHODS: RNA .....</b>	<b>103</b>
2.8.1	ISOLATION OF TOTAL RNA FROM CELL LINES .....	103
2.8.1.1	Using Trizol reagent .....	103
2.8.1.2	Using RNeasy kit .....	104
2.8.2	REMOVAL OF CONTAMINATING DNA FROM RNA .....	104
2.8.3	QUANTIFICATION OF RNA .....	105
2.8.4	REVERSE TRANSCRIPTION OF RNA .....	105
<b>2.9</b>	<b>MOLECULAR BIOLOGY METHODS: PROTEIN .....</b>	<b>106</b>
2.9.1	PROTEIN EXTRACTION FROM CELL LINES AND TISSUES .....	106
2.9.2	PREPARATION OF NUCLEAR AND CYTOPLASMIC EXTRACTS .....	106
2.9.3	SEPARATION OF PROTEINS ON POLYACRYLAMIDE GELS .....	106
2.9.4	WESTERN BLOTTING .....	107
2.9.5	ESTIMATION OF PROTEIN LOADING .....	108
<b>2.10</b>	<b>MOLECULAR BIOLOGY METHODS: POLYMERASE CHAIN REACTION .....</b>	<b>108</b>
2.10.1	STANDARD PCR .....	108
2.10.2	SEMIQUANTITATIVE PCR .....	109
2.10.3	NESTED PCR .....	109
2.10.4	PROOF-READING PCR .....	110
2.10.5	QUANTITATIVE PCR USING FLUORESCENT PROBES .....	110
2.10.6	USE OF TAQMAN LOW-DENSITY ARRAY PLATES .....	111
<b>2.11</b>	<b>MOLECULAR BIOLOGY METHODS: MUTATION DETECTION USING DHPLC (WAVE) .....</b>	<b>112</b>
2.11.1	PCR SET-UP FOR EXON AMPLIFICATION .....	112
2.11.2	WAVE HETERODUPLEX CONDITIONS .....	113
<b>2.12</b>	<b>MOLECULAR BIOLOGY METHODS: VECTORS .....</b>	<b>113</b>
2.12.1	INSERTION AND MANIPULATION OF VECTOR DNA .....	113
2.12.1.1	TOPO cloning .....	113
	Addition of 3' A overhangs to PCR products .....	114
2.12.1.2	DNA Ligation .....	114
	Blunt ended ligation .....	115
2.12.1.3	Site-directed mutagenesis .....	116
2.12.2	BACTERIAL TRANSFORMATION WITH VECTOR DNA .....	117
2.12.2.1	Transformation of bacterial strains .....	117
2.12.2.2	Selection strategies .....	118
	Blue-white colour screening .....	118
	Use of lethal genes .....	118
2.12.2.3	Isolation and purification of vectors from bacteria .....	118
	Small-scale isolation of plasmid DNA (miniprep) .....	118
	Large scale isolation of plasmid DNA (maxiprep) .....	119
2.12.3	TRANSFECTION OF VECTORS INTO MAMMALIAN CELL LINES .....	120
2.12.3.1	General considerations .....	120
2.12.3.2	Transfection with Fugene 6 reagent .....	121
2.12.3.3	Transfection with other reagents .....	121
2.12.3.4	Nucleoporation .....	122
2.12.3.5	Electroporation .....	122

2.12.4	ASSESSMENT OF EFFICIENCY OF TRANSFECTION .....	123
2.12.4.1	$\beta$ -galactosidase assay .....	123
	$\beta$ -Gal staining kit .....	123
	$\beta$ -Gal assay kit .....	123
2.12.4.2	Estimation of transfection efficiency using GFP .....	124
2.12.5	CELL SORTING FOR GFP EXPRESSION .....	124
<b>2.13</b>	<b>CELL BIOLOGY METHODS.....</b>	<b>125</b>
2.13.1	LUCIFERASE ASSAY .....	125
2.13.2	FLOW CYTOMETRY .....	125
2.13.2.1	Propidium iodide staining for DNA ploidy .....	126
2.13.3	IN VITRO DIFFERENTIATION WITH TPA .....	126
2.13.4	IN VITRO DIFFERENTIATION WITH HAEMIN .....	127
2.13.5	CYTOSPIN AND STAINING OF SUSPENSION CELLS .....	127
2.13.6	CONFOCAL MICROSCOPY .....	127
2.13.7	RNA INTERFERENCE.....	128
<b>2.14</b>	<b>EMBRYONIC STEM CELL METHODS.....</b>	<b>129</b>
2.14.1	SPECIAL CONSIDERATIONS FOR EMBRYONIC STEM CELL CULTURES.....	129
2.14.2	IN VITRO DIFFERENTIATION OF ES CELLS.....	129
2.14.2.1	Transcriptional profiling.....	129
2.14.2.2	Colony formation assays .....	130
2.14.3	PICKING AND STORING TARGETED ES CELL CLONES FOR SCREENING .....	131
2.14.3.1	Freezing and Thawing 24-well master plates .....	132
<b>2.15</b>	<b>STATISTICAL METHODS .....</b>	<b>132</b>
<b>3</b>	<b>AN ANALYSIS OF ENDOGENOUS GATA-1 ISOFORM EXPRESSION IN PRIMARY TISSUES AND CELL LINES.....</b>	<b>133</b>
<b>3.1</b>	<b>GATA-1 ISOFORMS ARE PRODUCED BY DIFFERENT MECHANISMS IN HUMANS AND MICE</b>	<b>134</b>
3.1.1	PCR PRIMERS SPANNING POTENTIAL SPLICE SITES SHOW ONE PRODUCT IN MICE AND TWO PRODUCTS IN HUMANS .....	134
3.1.2	WESTERN BLOTTING CONFIRMS PRESENCE OF 2 ISOFORMS IN MICE AND HUMANS .....	137
3.1.2.1	Optimisation of western blotting techniques .....	138
3.1.2.2	GATA-1 N6 antibody only detects murine GATA-1 .....	139
3.1.2.3	An intermittent additional band suggests variable post-translational modification.....	140
<b>3.2</b>	<b>PATTERNS OF GATA-1 ISOFORM EXPRESSION DURING HAEMATOPOIETIC DEVELOPMENT</b>	<b>140</b>
3.2.1	GATA-1 ISOFORM EXPRESSION IN HUMAN PRIMARY CELLS .....	141
3.2.2	GATA-1 EXPRESSION PATTERNS IN HUMAN CELL LINES .....	142
3.2.3	GATA-1 ISOFORM EXPRESSION DURING HUMAN FETAL DEVELOPMENT .....	145
3.2.4	GATA-1 ISOFORM EXPRESSION DURING INDUCTION OF TERMINAL ERYTHROID AND MEGAKARYOCYTIC DIFFERENTIATION .....	147
3.2.5	GATA-1 ISOFORM EXPRESSION DURING MURINE EMBRYOGENESIS .....	150
3.2.6	ISOFORM EXPRESSION IN MURINE CELL LINES.....	152
3.2.7	GATA-1 EXPRESSION DURING EMBRYONIC STEM CELL IN VITRO HAEMATOPOIETIC DIFFERENTIATION.....	154
<b>3.3</b>	<b>IDENTIFICATION OF ADDITIONAL GATA-1 TRANSCRIPTS .....</b>	<b>156</b>
3.3.1	IN SILICO ANALYSIS OF GATA-1 .....	156
3.3.1.1	GATA-1 exon analysis .....	156
3.3.1.2	GATA-1 EST analysis.....	157
3.3.1.3	GATA-1 related sequences held on the NCBI gene database .....	158
3.3.2	RT-PCR ANALYSIS OF MRNA .....	159
3.3.3	WESTERN BLOTTING REVEALS ONLY TWO BANDS .....	160
<b>3.4</b>	<b>IN VITRO ANALYSIS OF BC009797.....</b>	<b>161</b>
3.4.1	ANALYSIS OF IMAGE CLONE REASSIGNS CLONING VECTOR BUT CONFIRMS SEQUENCE ..	161
3.4.2	ANALYSIS OF BC009797 EXPRESSION AT THE MRNA LEVEL .....	163
3.4.3	A FUNCTIONAL ANALYSIS OF BC009797.....	164
3.4.3.1	Western blotting is unable to detect BC009797 .....	165



3.4.3.2	Reporter assays suggest BC009797 has enhanced transactivation potential .....	166
3.4.4	SUMMARY OF THE RELEVANCE OF BC009797 .....	167
<b>3.5</b>	<b>CHAPTER 3 SUMMARY.....</b>	<b>168</b>
<b>4</b>	<b>GATA-1 GENE TARGETING IN MURINE EMBRYONIC STEM CELLS.....</b>	<b>170</b>
<b>4.1</b>	<b>AIMS.....</b>	<b>170</b>
<b>4.2</b>	<b>INTRODUCTION.....</b>	<b>170</b>
4.2.1	TARGETING SELECTION STRATEGIES .....	173
4.2.1.1	Positive selection .....	173
4.2.1.2	Negative selection.....	173
4.2.2	OPTIMAL DESIGN OF THE TARGETING VECTOR.....	174
4.2.2.1	Homology arms .....	174
4.2.2.2	Removal of the selectable marker.....	174
4.2.3	METHODS TO SCREEN FOR RECOMBINATION EVENTS.....	175
4.2.3.1	PCR.....	176
4.2.3.2	Southern blotting .....	178
<b>4.3</b>	<b>CREATION OF GATA-1 ISOFORM TARGETING CONSTRUCTS.....</b>	<b>180</b>
4.3.1	GENERATING HOMOLOGY ARMS.....	180
4.3.2	INTRODUCTION OF POINT MUTATIONS INTO FIRST OR SECOND START CODONS .....	180
4.3.3	ADDITION OF SELECTABLE MARKERS.....	182
4.3.4	DEVISING AND TESTING A SCREENING STRATEGY FOR TARGETED CLONES .....	182
<b>4.4</b>	<b>TARGETING CONSTRUCT VERSION 1.....</b>	<b>183</b>
4.4.1	VECTOR A PCR4 EXON3-6.....	187
4.4.1.1	Site-directed mutagenesis of Vector A .....	189
4.4.2	VECTOR B PCR4 EXON2 .....	190
<b>4.5</b>	<b>TARGETING CONSTRUCT VERSION 2.....</b>	<b>194</b>
4.5.1	STEP 1 CLONING OF VECTOR 1 (PBSII EXON 2).....	197
4.5.1.1	Site-directed mutagenesis of vector 1 .....	197
4.5.2	STEP 2 CLONING OF VECTOR 2 (PBSII EXON2-EXON3-6).....	198
4.5.3	STEP 3 INTRODUCTION OF PGKNEO.....	199
4.5.4	STEP 4 INTRODUCTION OF DIPHTHERIA TOXIN .....	200
<b>4.6</b>	<b>INTRODUCTION INTO ES CELLS .....</b>	<b>202</b>
<b>4.7</b>	<b>SCREENING COLONIES .....</b>	<b>202</b>
4.7.1	PCR SCREENING .....	203
4.7.1.1	Optimisation of technique .....	203
4.7.1.2	Results of attempt 1 .....	205
4.7.2	SOUTHERN BLOTTING .....	207
4.7.2.1	Preparation of 3' and 5' probes .....	207
4.7.2.2	Results of Southern blotting .....	207
4.7.3	INCREASING THE SENSITIVITY OF PCR SCREENING.....	208
<b>4.8</b>	<b>SECOND AND THIRD TARGETING ATTEMPTS .....</b>	<b>209</b>
4.8.1	SCREENING BY NESTED PCR .....	210
4.8.2	SCREENING BY SOUTHERN BLOT .....	212
4.8.3	FURTHER INVESTIGATION OF POSITIVE CLONES .....	212
4.8.3.1	Are the PCR bands false positives? .....	216
4.8.3.2	Is there any evidence for mixed colonies?.....	216
4.8.3.3	Do clones with a targeted band also possess wild-type GATA-1?.....	218
<b>4.9</b>	<b>SUMMARY OF TARGETING EXPERIMENTS .....</b>	<b>219</b>
<b>5</b>	<b>A FUNCTIONAL ANALYSIS OF GATA-1 ISOFORMS.....</b>	<b>221</b>
<b>5.1</b>	<b>INTRODUCTION.....</b>	<b>221</b>
<b>5.2</b>	<b>CONSTRUCTION OF GATA-1 ISOFORM SPECIFIC EXPRESSION VECTORS .....</b>	<b>222</b>
5.2.1	HUMAN GATA-1 EXPRESSION VECTORS.....	222
5.2.2	MURINE GATA-1 EXPRESSION VECTORS .....	224
5.2.3	CONSTRUCTION OF pCDNAGATA-1FL, pCDNAGATA-1S AND pCDNAGATA-1WT ...	224
5.2.4	VALIDATION OF pCDNAGATA-1 EXPRESSION VECTORS .....	226

5.2.4.1	Expression at the mRNA level .....	226
5.2.4.2	Expression at the protein level.....	226
5.2.4.3	Production of functional protein.....	228
5.2.5	DESIGN OF ADDITIONAL EXPRESSION VECTORS .....	231
5.2.5.1	EF1 $\alpha$ vectors.....	231
5.2.5.2	IRES-GFP vectors .....	233
<b>5.3</b>	<b>EXPRESSION OF GATA-1 ISOFORMS IN MURINE ES CELLS.....</b>	<b>235</b>
5.3.1	DETECTION OF GATA-1 EXPRESSION.....	235
5.3.2	EFFECTS OF GATA-1 ISOFORM EXPRESSION VECTORS ON HAEMATOPOIETIC DIFFERENTIATION OF ES CELLS .....	239
<b>5.4</b>	<b>EXPRESSION OF GATA-1 ISOFORMS IN K562 AND MEG-01 CELLS – CREATION OF TRANSGENIC LINES AND STUDY OF BASELINE CHARACTERISTICS .....</b>	<b>241</b>
5.4.1	STABLE TRANSFECTION .....	241
5.4.2	SELECTION FOR HIGH TRANSGENE EXPRESSION .....	243
5.4.3	SPONTANEOUS SELECTION FOR ISOFORM-SPECIFIC TRANSGENE EXPRESSION ON CONTINUED CELL CULTURE .....	243
5.4.4	METHODS OF EVALUATION OF BASELINE CHARACTERISTICS.....	244
5.4.5	CELL GROWTH RATES .....	245
5.4.6	MATURATION STATUS OF CELLS.....	247
5.4.6.1	Evaluation of cell morphology .....	247
5.4.6.2	Measurement of DNA ploidy .....	249
5.4.6.3	Analysis of cell surface marker expression .....	250
5.4.7	GENE EXPRESSION PROFILES OF TRANSFECTED CELL LINES.....	252
5.4.7.1	K562 vs. Meg-01 gene expression.....	256
5.4.7.2	The effects of GATA-1 transgene expression in K562 cells .....	259
5.4.7.3	The effects of GATA-1 transgene expression in Meg-01 cells .....	262
<b>5.5</b>	<b>THE USE OF K562 AND MEG-01 CELLS TO MODEL MEGAKARYOCYTIC AND ERYTHROID DIFFERENTIATION.....</b>	<b>265</b>
5.5.1	VALIDATION OF MODEL SYSTEMS .....	266
5.5.1.1	TPA.....	266
5.5.1.2	Haemin .....	271
5.5.2	THE EFFECTS OF ISOFORM EXPRESSION ON RESPONSE TO TPA IN K562 AND MEG-01 CELLS 273	
5.5.2.1	Effects on DNA ploidy .....	273
5.5.2.2	Effects on gene expression profiles .....	275
5.5.2.3	Gene expression in K562 transgenics.....	276
5.5.2.4	Gene expression in Meg-01 transgenics.....	279
5.5.3	THE EFFECTS OF ISOFORM EXPRESSION ON RESPONSE TO HAEMIN IN K562 AND MEG-01 CELLS 281	
5.5.3.1	Haemoglobin production in K562 cells.....	281
5.5.3.2	Gene expression in K562 transgenics.....	282
5.5.3.3	Gene expression in Meg-01 transgenics.....	284
5.5.4	THE IMPACT OF GATA-1 ISOFORM EXPRESSION ON ERG3 AND BACH-1 .....	286
5.5.5	ALTERNATIVE APPROACHES TO STUDYING THE EFFECTS OF ISOFORM EXPRESSION .....	288
5.5.6	SIRNA MEDIATED KNOCKDOWN OF HUMAN GATA-1 .....	289
5.5.6.1	Optimisation of transfection efficiency .....	291
5.5.6.2	Effects of siRNA mediated GATA-1 targeting .....	292
5.5.6.3	Discussion of siRNA results.....	294
<b>5.6</b>	<b>SUMMARY OF THE RESULTS OF FUNCTIONAL ASSESSMENT OF GATA-1 ISOFORMS IN MURINE AND HUMAN CELL LINES .....</b>	<b>296</b>
<b>6</b>	<b>ANALYSIS OF GATA-1 MUTATIONAL STATUS IN CHRONIC MYELOID LEUKAEMIA .....</b>	<b>298</b>
<b>6.1</b>	<b>BACKGROUND.....</b>	<b>298</b>
<b>6.2</b>	<b>IN SILICO ANALYSIS.....</b>	<b>299</b>
<b>6.3</b>	<b>MUTATIONAL ANALYSIS USING PATIENT SAMPLES.....</b>	<b>299</b>
6.3.1	SAMPLES .....	299

6.3.2	OPTIMISATION OF METHODS TO DETECT GATA-1 MUTATIONS .....	300
6.3.3	APPROACH 1: PCR OF CODING EXONS FOLLOWED BY DIRECT SEQUENCING OF PRODUCTS 300	
6.3.4	APPROACH 2: PCR OF CODING EXONS FOLLOWED BY ANALYSIS OF PRODUCTS BY DHPLC 302	
6.3.4.1	Principles of WAVE technique .....	302
6.3.4.2	WAVE analysis of CML patient samples.....	304
6.3.4.3	Sequence analysis of samples with abnormal products .....	306
6.3.5	VERIFICATION OF RESULTS USING TOPO CLONING OF PCR PRODUCTS .....	307
<b>6.4</b>	<b>SUMMARY .....</b>	<b>307</b>
<b>7</b>	<b>CONCLUSIONS AND FUTURE DIRECTIONS.....</b>	<b>309</b>
<b>7.1</b>	<b>SUMMARY OF FINDINGS .....</b>	<b>309</b>
<b>7.2</b>	<b>HOW DO THESE FINDINGS EXTEND CURRENT KNOWLEDGE OF GATA-1 FUNCTION?.....</b>	<b>310</b>
7.2.1	THE PHYSIOLOGICAL ROLE OF GATA-1S .....	310
7.2.2	THE ROLE OF GATA-1 IN TAM/AMKL .....	316
7.2.3	THE ROLE OF GATA-1 IN OTHER MALIGNANCIES .....	319
<b>7.3</b>	<b>FUTURE INVESTIGATIONS.....</b>	<b>321</b>
<b>8</b>	<b>REFERENCES .....</b>	<b>325</b>

# List of Tables

---

## **1: Introduction**

Table 1.1: Classification of transcription factors .....	33
Table 1.2: Master Haematopoietic Transcription Factors.....	35

## **2: Materials and Methods**

Table 2.1: Solution composition .....	84
Table 2.2 Cell lines and culture requirements.....	87
Table 2.3: Bacterial culture media .....	88
Table 2.4: GATA-1 primers .....	88
Table 2.5: Primers used for vector construction .....	89
Table 2.6: Primers for site directed mutagenesis .....	89
Table 2.7: Primers for RT-PCR transcriptional profiling .....	90
Table 2.8: CML direct sequencing primers .....	90
Table 2.9: CML WAVE primers.....	90
Table 2.10: ES cell gene targeting primers .....	91
Table 2.11: qPCR primer/probe sets .....	91
Table 2.12: RNAi duplexes.....	91
Table 2.13: Antibodies for Western blotting .....	92
Table 2.14: Antibodies for flow cytometry.....	92

## **3: An analysis of endogenous GATA-1 isoform expression in primary tissues and cell lines**

Table 3.1 Human haematopoietic cell lines and their normal counterparts.....	143
Table 3.2 Murine cell lines .....	152

## **4: GATA-1 gene targeting in murine embryonic stem cells**

Table 4.1 Investigation of possible targeted clones .....	214
---	-----

## **5: A functional analysis of GATA-1 isoforms**

Table 5.1 Genes included in qPCR expression assays.....	254
---	-----

# List of Figures

---

## **1: Introduction**

Fig 1.1 Hematopoietic differentiation.....	22
Fig 1.2: Sites and types of haematopoiesis during fetal life .....	25
Fig 1.3. The epigenetic control of gene expression.....	38
Fig 1.4 Mechanisms of gene regulation by micro-RNAs .....	39
Fig 1.5 GATA-1 Structure .....	42
Fig 1.6 GATA-1 autoregulation and cross regulation with GATA-2 during commitment to the erythroid lineage.....	50
Fig 1.7 At least 3 hits are needed to develop AMKL .....	67
Fig 1.8 Proposed impact of clone size and acquisition of mutations on clinical presentation .....	70
Fig 1.9 Proposed model of combinatorial effects of chromosome 21 gene overexpression and additional mutations on development of AMKL. ....	76

## **2: Materials and Methods**

Fig 2.1 Assembly of Southern blot downward transfer apparatus .....	101
Fig 2.2 in vitro differentiation of ES cells .....	130

## **3: An analysis of endogenous GATA-1 isoform expression in primary tissues and cell lines**

Fig 3.1 Detection of alternative splicing by PCR .....	135
Fig 3.2 PCR does not reveal alternative splicing in mice .....	136
Fig 3.3 Optimisation of western blotting technique.....	138
Fig 3.4 Characteristics of available antibodies .....	139
Fig 3.5 GATA-1 isoform expression in human primary cells .....	142
Fig 3.6 GATA-1 expression in human cell lines .....	144
Fig 3.7 GATA-1 isoform expression in human embryonic tissues.....	146
Fig 3.8 Variation of GATA-1 isoforms during in vitro erythroid differentiation.....	148
Fig 3.9 GATA-1 isoform expression during in vitro megakaryocytic differentiation of cell lines .....	149
Fig 3.10 GATA-1 isoform expression during mouse embryonic development.....	151
Fig 3.11 GATA-1 isoform expression in murine haemopoietic cell lines .....	153
Fig 3.12 GATA-1 expression during in vitro haematopoietic differentiation of murine embryonic stem cells.....	155
Fig 3.13 Generation of expressed sequence tags .....	157
Fig 3.14 Comparison between NM_002049 and BC009797 .....	158
Fig 3.15 PCR strategies to detect novel GATA-1 transcripts .....	159
Fig 3.16 Detection and analysis of an additional 1500bp transcript in human samples ....	160
Fig 3.17 Restriction enzyme analysis of IMAGE clone suggests an alternative cloning vector.....	162
Fig 3.18 PCR screening for BC009797-like transcript in other human primary cells.....	164
Fig 3.19 BC009797 mRNA and protein expression .....	165
Fig 3.20 Function of BC009797 in a luciferase reporter assay.....	166

## **4: GATA-1 gene targeting in murine embryonic stem cells**

Fig 4.1 Outline of gene targeting approach.....	172
Fig 4.2 Cre Lox-P mediated recombination to excise the selectable marker.....	176

Fig 4.3 Principles of PCR screening for targeted colonies .....	177
Fig 4.4 Principles of Southern blot analysis for targeted colonies.....	179
Fig 4.5 Production of isoform specific targeting vectors.....	181
Fig 4.6 Targeting construct version 1 .....	184
Fig 4.7 Screening strategy for targeted colonies.....	186
Fig 4.8 Restriction enzyme digestion of Vector A.....	188
Fig 4.9 Site-directed mutagenesis of Vector A .....	189
Fig 4.10 Optimisation of exon 2 amplification .....	191
Fig 4.11 Restriction enzyme digestion of clone 11 vector B.....	192
Fig 4.12 Attempt to reverse orientation of clone 11 .....	193
Fig 4.13 Outline of targeting strategy 2 .....	196
Fig 4.14 Site-directed mutagenesis of vector 1 .....	198
Fig 4.15 Diagnostic digests to confirm construction of the targeting vector.....	201
Fig 4.16 Construction of PCR+ve control plasmid.....	203
Fig 4.17 Titration of template DNA to confirm utility of PCR screening .....	205
Fig 4.18 PCR screening of targeting attempt 1 .....	206
Fig 4.19 Southern blotting of targeting attempt 1 .....	208
Fig 4.20 Nested PCR screening strategy.....	209
Fig 4.21 Nested PCR results for targeting attempts 1 and 2 .....	210
Fig 4.22 Positive PCR screening results from T2 and T3.....	211
Fig 4.23 Southern blotting of targeting attempts 2 and 3.....	213
Fig 4.24 Screening of FL40 subclones by PCR and Southern blot.....	217
Fig 4.25 PCR reveals wild-type GATA in all colonies tested .....	218

## **5: A functional analysis of GATA-1 isoforms**

Fig 5.1 Assessment of Human GATA-1 expression plasmids.....	223
Fig 5.2 Creation of GATA-1 isoform specific products .....	225
Fig 5.3 Construction and expression from pcDNA vectors .....	227
Fig 5.4 Luciferase assay of GATA-1 transactivation potential .....	230
Fig 5.5 Cloning into the pEF6 expression vector and addition of IRES-GFP .....	232
Fig 5.6 Expression of GFP in transfected cell lines.....	234
Fig 5.7 Detection of transgene expression in murine ES cells.....	238
Fig 5.8 Effects of GATA-1 isoform expression on in vitro differentiation of murine ES cells .....	240
Fig 5.9 Stable expression of GATA-1 isoforms in K562 and Meg-01 cells.....	242
Fig 5.10 K562 and Meg-01 growth curves .....	246
Fig 5.11 Influence of transgene expression on cell morphology .....	248
Fig 5.12 Ploidy analysis of K562 and Meg-01 stable transfectants.....	249
Fig 5.13 Expression of erythroid and megakaryocytic cell surface markers in Meg-01 and K562 transgenic cell lines .....	251
Fig 5.14 Comparison of gene expression between Meg-01 and K562 .....	258
Fig 5.15 Alterations in gene expression profile with GATA-1 isoform over-expression in K562 cells .....	260
Fig 5.16 Alterations in gene expression profile with GATA-1 isoform over-expression in Meg-01 cells.....	262
Fig 5.17 RT-PCR for megakaryocytic and erythroid genes following TPA treatment of Meg-01 and K562 cell lines .....	267
Fig 5.18 The effect of TPA treatment on DNA ploidy and cell surface marker expression .....	269
Fig 5.19 Effects of haemin treatment.....	272
Fig 5.20 DNA ploidy changes in response to TPA .....	274
Fig 5.21 Variation of 18SrRNA levels with TPA and Haemin treatment .....	276
Fig 5.22 Gene expression patterns in TPA treated K562 transgenic cells.....	277

Fig 5.23 Gene expression patterns in TPA treated Meg-01 transgenic cells .....	280
Fig 5.24 Gene expression changes in K562 transgenics treated with Haemin .....	283
Fig 5.25 Gene expression changes in Meg-01 transgenics treated with haemin .....	285
Fig 5.26 Erg3 and Bach-1 expression levels in K562 and Meg-01 cells .....	287
Fig 5.27 Estimation of siRNA transfection efficiency using Rhodamine labelled RISC free siRNA.....	291
Fig 5.28 siRNA knockdown of GATA-1 at the protein level.....	292
Fig 5.29 Results of qPCR analysis of siRNA mediated GATA-1 knockdown.....	294

## **6: Analysis of GATA-1 mutational status in chronic myeloid leukaemia**

Fig 6.1 EST sequence error rate in CML and normal individuals .....	299
Fig 6.2 PCR of coding exons with direct sequencing of products .....	301
Fig 6.3 Titration of genomic DNA template for high and low efficiency PCR reactions ..	302
Fig 6.4 Principles of WAVE mutation screening using DHPLC.....	303
Fig 6.5 Amplification of coding exons of GATA-1 for analysis by DHPLC .....	304
Fig 6.6 Chromatogram traces from WAVE mutational analysis .....	305
Fig 6.7 WAVE and sequencing chromatograms for the exon 5 fragment of samples M949BC and M84BC compared to wild-type DNA. ....	306
Fig 7.1 A potential role for GATA-1s in haematopoietic progenitor cells .....	313
Fig 7.2 High levels of GATA-1s may be able to compensate for loss of GATA-1FL .....	314

## List of publications arising from this work

---

1. **C Halsey**, C Fisher, G Strathdee, B Gibson, T Holyoake, P Vyas, G Graham. GATA-1 mutational analysis in chronic myeloid leukaemia. *British Journal of Haematology* (2007), **137**, p375-376
  
2. C McKimmie, A Fraser, C Hansell, L Gutierrez, S Philipsen, L Connell, A Rot, M Kurowska-Stolarska, P Carreno, M Pruenster, C Chu, G Lombardi, **C Halsey**, I McInnes, F Liew, R Nibbs, G Graham. Haemopoietic cell expression of the chemokine decoy receptor D6 is dynamic and regulated by GATA-1. *Journal of Immunology* (2008), **181**, p3353-3363.



## Acknowledgements

---

I am extremely grateful to my supervisors Professor Gerry Graham and Dr Brenda Gibson and all the members of the Chemokine Research Group for their support, practical help and encouragement over the last few years. In addition, I would like to specifically acknowledge the invaluable help of Dr Derek Gilchrist in devising and implementing gene targeting strategies and for sharing his in depth knowledge of molecular biology. A number of other individuals also assisted with setting up of specific techniques including; Dr Clive McKimmie (qPCR), Dr Alasdair Fraser and Dr Chris Hansell (FACS), Dr Ashley Gilmour (FACS sorting), and Claire Burt (confocal microscopy). I would also like to thank Marie-Helena Docherty and Mhairi McNeill for their input and enthusiasm. A number of collaborators also assisted with samples and ideas including Dr Pares Vyas, Dr Chris Fisher (WAVE, Luciferase reporter vector) and Dr Isla Hamlett (western blot nuclear extraction protocol) from the Weatherall Institute of Molecular Medicine, Oxford and Dr Gordon Strathdee (CML genomic DNA) from the Beatson Institute for Cancer Research, Glasgow. I am also very grateful to Professor Irene Roberts and Dr Oliver Tunstall-Pedoe (Imperial College, London) for their sharing of human embryonic samples and data.

Finally, I gratefully acknowledge the generous support of the Leukaemia Research Fund, whose award of an LRF Clinical Training Fellowship enabled me to carry out this research.

## Declaration

I declare that I am the sole author of this thesis and that all the work described within was performed by me unless otherwise stated.

## Abbreviations

---

<b>AGM</b>	Aorta gonad mesonephros
<b>ALL</b>	Acute lymphoblastic leukaemia
<b>AMKL</b>	Acute megakaryoblastic leukaemia
<b>AML</b>	Acute myeloid leukaemia
<b>BFU-E</b>	Burst forming unit – erythroid
<b>BM</b>	Bone marrow
<b>Bp</b>	Base-pair
<b>cDNA</b>	Complementary deoxyribonucleic acid
<b>CFC</b>	Colony forming cell
<b>CFU-GM</b>	Colony forming unit – granulocyte macrophage
<b>ChIP</b>	Chromatin immunoprecipitation
<b>CLP</b>	Common lymphoid progenitor
<b>CML</b>	Chronic myeloid leukaemia
<b>CMP</b>	Common myeloid progenitor
<b>DC</b>	Dendritic cell
<b>ddH<sub>2</sub>O</b>	Double distilled water
<b>DHPLC</b>	Denaturing high-performance liquid chromatography
<b>DNA</b>	Deoxyribonucleic acid
<b>DS</b>	Down syndrome
<b>E</b>	Days post-coitus
<b>EB</b>	Embryoid body
<b>EDTA</b>	Ethyl diamine tetra acetate
<b>EKLF</b>	Erythroid Krüppel-like factor
<b>EPO</b>	Erythropoietin
<b>ES cells</b>	Embryonic stem cells
<b>EST</b>	Expressed sequence tag
<b>FACS</b>	Fluorescent activated cell sorting
<b>FOG-1</b>	Friend of GATA-1
<b>FSC</b>	Forward Scatter
<b>GATA-1FL</b>	Full-length GATA-1 isoform
<b>GATA-1s</b>	Short GATA-1 isoform
<b>GFP</b>	Green fluorescent protein
<b>GMCSF</b>	Granulocyte macrophage colony stimulating factor
<b>HAT</b>	Histone acetyltransferase
<b>HDAC</b>	Histone deacetylase
<b>HIAC</b>	Haematopoietic intra-aortic cluster

<b>HIS</b>	Histidine
<b>HMT</b>	Histone methyltransferase
<b>HS</b>	DNase I hypersensitive site
<b>HRP</b>	Horseradish peroxidase
<b>HSC</b>	Haematopoietic stem cell
<b>IE</b>	GATA-1 erythroid first exon
<b>IRES</b>	Internal ribosome entry site
<b>IT</b>	GATA-1 testis first exon
<b>JAK</b>	Janus activating kinase
<b>kB</b>	Kilobase
<b>kDa</b>	KiloDalton
<b>LIF</b>	Leukaemia inhibitory factor
<b>LTR</b>	Long terminal repeat
<b>LTR-HSC</b>	Long-term repopulating haematopoietic stem cell
<b>MEL</b>	Mouse erythroleukaemia
<b>MEP</b>	Megakaryocyte-erythroid progenitor
<b>MPO</b>	Myeloperoxidase
<b>MPP</b>	Multipotent progenitor
<b>mRNA</b>	Messenger ribonucleic acid
<b>miRNA</b>	Micro ribonucleic acid
<b>MW</b>	Molecular weight
<b>PBS</b>	Phosphate buffered saline
<b>PF4</b>	Platelet Factor 4
<b>PCR</b>	Polymerase chain reaction
<b>Pol II</b>	RNA polymerase II
<b>qPCR</b>	Quantitative polymerase chain reaction
<b>RE</b>	Restriction enzyme
<b>RQ</b>	Relative quantitation
<b>RNA</b>	Ribonucleic acid
<b>RNAi</b>	RNA interference
<b>RT-PCR</b>	Reverse transcriptase polymerase chain reaction
<b>SAP</b>	Sub-aortic patch
<b>siRNA</b>	Small interfering ribonucleic acid
<b>SSC</b>	Side scatter
<b>STAT</b>	Signal transducer and activator of transcription
<b>ST-HSC</b>	Short-term haematopoietic stem cell
<b>TAM</b>	Transient abnormal myelopoiesis
<b>UV</b>	Ultraviolet

# 1 Introduction

---

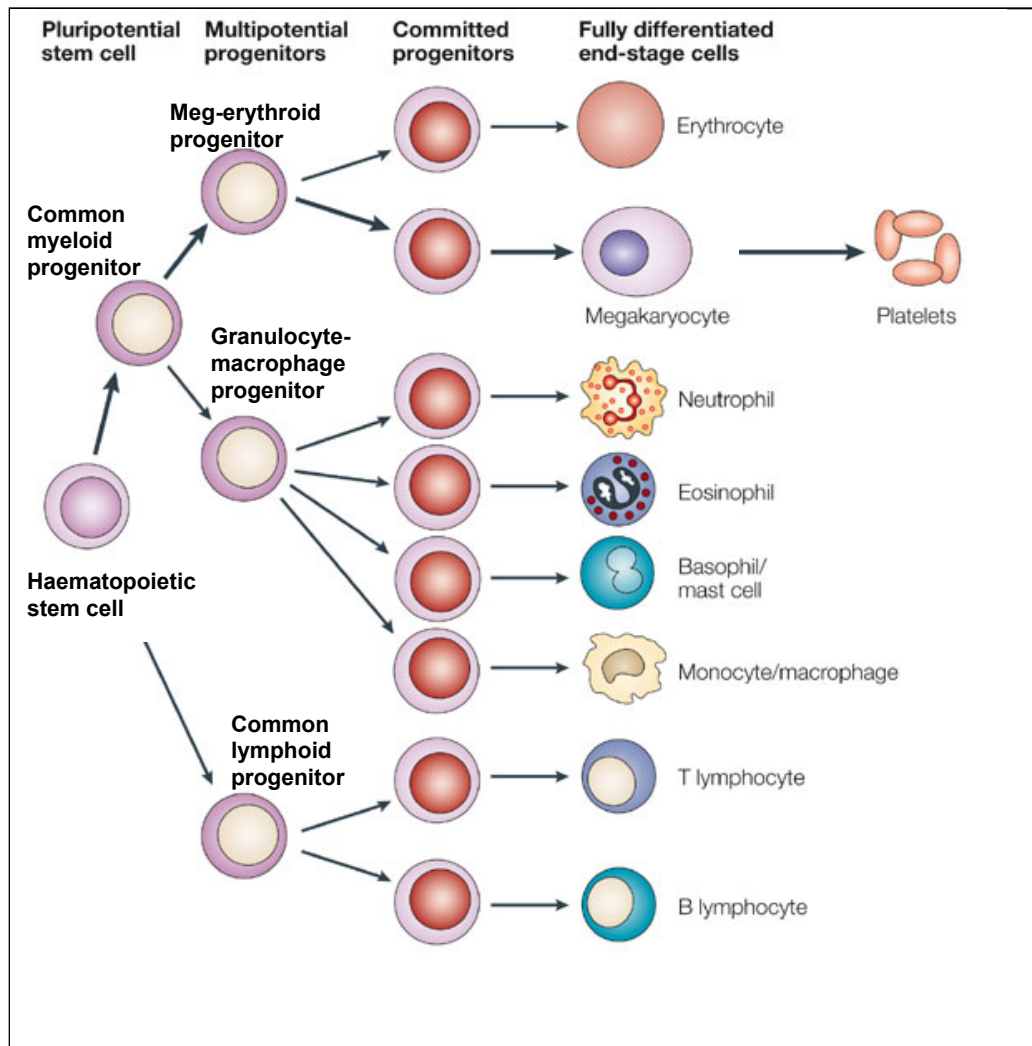
All cells in the body contain essentially identical genetic information and yet they can differentiate into a wide variety of cell types each with a unique function. The mechanisms underlying these choices to become one cell type or another are only partially understood and have many layers of complexity. In general, differentiation involves expression of a set of genes needed for specialised cellular functions and repression of genes not needed in that cell type. Transcription factors play a key role in this process, these are proteins that bind DNA upstream of genes and help determine whether the gene is expressed or not. In addition, tissues capable of self-renewal must possess a subset of cells known as stem cells. Understanding the mechanisms by which cells self-renew and differentiate is vital to both basic biology and many aspects of health and disease including the formation of cancers, ageing and developmental anomalies.

Haematopoiesis describes the generation of large numbers of mature blood cells from small numbers of haematopoietic stem cells. In this introduction the cellular and molecular basis of mammalian haematopoietic cell differentiation will be discussed, with particular reference to the critical role of transcription factors. A key haematopoietic transcription factor – GATA-1- forms the basis of this thesis and will be discussed in depth. The introduction will end by addressing the relevance of this basic science to understanding the molecular pathogenesis of certain leukaemias by focussing on a rare but highly informative leukaemia – acute megakaryoblastic leukaemia associated with trisomy 21 (Down syndrome).

## 1.1 Haematopoiesis

Blood cells comprise at least eight mature cell types with different functions and lifespan. Most only survive a few hours (granulocytes) to a few weeks (erythrocytes). Every day approximately  $10^{13}$  new myeloid cells are required to replace these losses (Hoffbrand, *et al* 2005). In addition, certain subsets are capable of rapid change in response to external stresses such as hypoxia, infection and inflammation. Within the adult bone marrow, blood cells are found in various stages of development. It has long been known that bone marrow cells can regenerate the blood compartment, but the idea of multipotent progenitors capable of producing a number of different blood cell types emerged in the 1960s and 70s following seminal work by Metcalf and Moore and by Till and McCullough. These groups

documented the existence of colony forming units (CFUs), derived from mouse bone marrow, capable of producing cells of different lineages when transferred to irradiated recipients and then subdivided them on the basis of their response to various colony stimulating factors (Metcalf 1970a, Metcalf 1999, Till and McCulloch 1961). At the top of the haematopoietic progenitor tree sit haematopoietic stem cells (HSCs). These stem cells comprise about 0.01-0.05% of the total marrow population, hence the production of mature cells from these relatively rare stem cells involves both expansion and differentiation. The stages of cellular differentiation are shown in Fig 1.1.



**Fig 1.1 Hematopoietic differentiation.**

Adapted from Hitzler and Zipursky, *Nature Reviews Cancer* (Hitzler and Zipursky 2005)

The sites of haematopoiesis and the exact nature of the cells produced vary during development. This reflects the changing needs of the embryo, fetus and adult organism. After birth the majority of haematopoiesis in the resting state occurs in the bone marrow. However before birth the fetal liver is the main haematopoietic organ. In order to

understand the molecular and cellular events underlying haematopoiesis it is important to be aware of changes in the system during ontogeny.

### **1.1.1 Haematopoiesis during development**

Since blood cells are essential for carriage of oxygen to the developing cells in the early embryo, haematopoiesis and accompanying vasculogenesis (formation of blood vessels) are very early events in embryogenesis and major defects in the genes critical for blood formation are usually lethal at an early stage. Haematopoietic cells arise from primitive embryonic mesoderm. The process by which mesoderm is induced to become haematopoietic tissue is largely unknown, although the gene Bone Morphogenetic Protein 4 (BMP-4) appears to play a particularly important role (Johansson and Wiles 1995). Blood cell development occurs in at least two phases. The first occurs within the blood islands of the yolk-sac – known as primitive (embryonic) haematopoiesis and the second arises within the embryo itself and produces mature haematopoietic stem cells capable of multi-lineage blood cell production - known as definitive (adult) haematopoiesis. The relationship between the development of the vascular and haematopoietic systems remains controversial although evidence is now accumulating in favour of a shared progenitor cell known as the haemangioblast (Huber, *et al* 2004). It is important to be aware of the differences between embryonic and adult blood production when considering the effects of mutations in haematopoietic genes as these may vary depending on the developmental stage at which the mutation was acquired and/or the stage at which the gene is normally active.

#### **1.1.1.1 Extra-embryonic and intra-embryonic haematopoiesis**

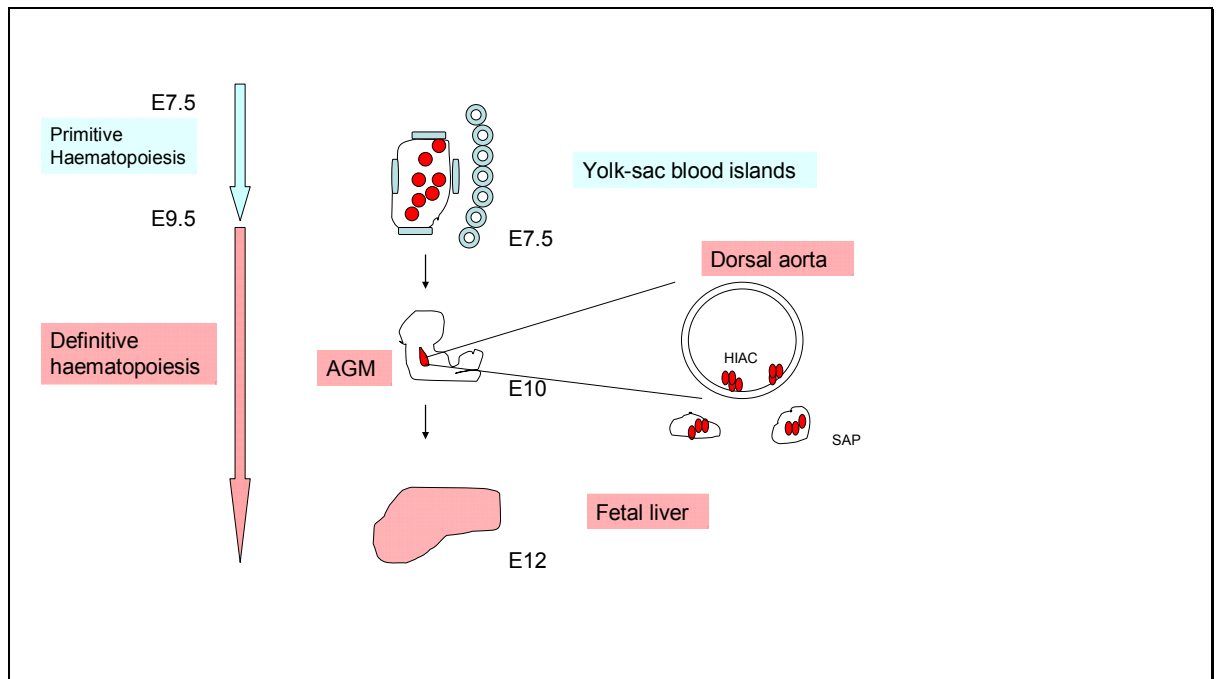
The first recognisable blood cells are seen in the extra-embryonic yolk sac at 7-7.5 days post coitus / E7-7.5 (days post coitus is referred to as the prefix E) in the mouse embryo (Cumano and Godin 2007). Mesodermal cells initially form homogenous aggregates known as haemangioblastic chords. They then form discrete structures known as blood islands comprising a central cluster of blood cells surrounded by cells with endothelial characteristics. The blood cells appear to be mainly erythrocytes which are large and nucleated and resemble both the early proerythroblasts seen in adult bone marrow and erythrocytes from lower vertebrates such as fish and birds. For these reasons they are known as primitive erythrocytes. Later in embryogenesis blood cell formation shifts to the fetal liver and then the thymus, spleen and bone marrow. Initially it was thought this represented sequential migration to, and colonisation of, these organs by the HSCs that had

initially formed in the yolk-sac. Early experiments using transplantation assays and formation of interspecies chimeras (such as quail-chick grafts) established that, although the yolk-sac was capable of a partial contribution to haematopoiesis after birth, there appeared to be separate areas capable of generation of haematopoietic stem cells within the embryo itself (Dieterlen-Livre 1975).

The intra-embryonic site of haematopoiesis was narrowed down by a combination of grafting experiments (Godin, *et al* 1993) and *in vitro* culture and cytological analysis (Tavian, *et al* 1999). Haematopoietic capability appeared to be located in the dorsal aorta and its underlying mesenchyme. This aortic region is also referred to as the Aorta-Gonads-Mesonephros region (AGM), describing its developmental fate. Generation of HSCs in this region occurs at E10.5 (Medvinsky and Dzierzak 1996), establishment of a circulation between the yolk-sac and embryo itself occurs at E8. This left open the question as to whether these areas were capable of independent HSC generation or whether they merely provided a permissive environment for colonisation and expansion of yolk-sac derived HSCs. Elegant organ explant cultures of intact para-aortic splanchnopleura (the mesodermal-endodermal association created by subdivision of the lateral plate mesoderm by the formation of the coelomic cavity early in development) isolated from the day 8 embryo, before establishment of a yolk-sac derived circulation, have proved that this region is indeed capable of independent generation of HSCs (Cumano, *et al* 1996). Further work has identified the precise locations within the AGM important for generation of intraembryonic HSCs. Firstly, the existence of haematopoietic intra-aortic clusters (HIACs) has been demonstrated; comprising immature haematopoietic cells as well as some more differentiated cells. These are located in the ventral wall of the dorsal aorta intimately associated with endothelial cells. Secondly, a region below the aortic floor known as the subaortic patches (SAPs) was identified due to its strong expression of genes and proteins important in haematopoietic development. Its appearance precedes that of HIACS leading to the hypothesis that HSCs may arise there and migrate to the HIAC region (Cumano and Godin 2007). Fig 1.2 summarises the important features of intra and extra-embryonic haematopoiesis discussed above. Besides differences in location, intra-embryonic HSCs appear functionally different from yolk-sac HSCs. They have much more extended differentiation capacity being capable of formation of lymphoid and myeloid cells whilst yolk sac HSCs are unable to contribute to lymphopoiesis in transplant experiments (Yokota, *et al* 2006). Although the majority of experiments described above have been carried out using avian and murine models, most of these observations have been confirmed in humans (Tavian, *et al* 2001).



The debates as to the origin of HSCs in yolk sac versus dorsal aorta have recently acquired new relevance with the unexpected observation that the placenta, another extra-embryonic site, may provide an additional source of HSCs in the mouse (Corbel, *et al* 2007, Zeigler, *et al* 2006). At E11 (at the same time as fetal liver colonisation from dorsal aorta derived HSCs), cells recovered from the placenta show long-term repopulating (LTR) activity in transplantation assays (a classic assay for HSC activity) (Gekas, *et al* 2005, Ottersbach and Dzierzak 2005). Three possibilities exist to explain their presence – firstly they may represent circulating HSCs generated in the dorsal aorta and preferentially isolated from the placenta due to its large blood supply. Secondly the placenta may provide an appropriate niche for the expansion and survival of HSCs generated elsewhere. Finally the placenta may be a site of *de novo* generation of HSCs. Overall the weight of experimental evidence probably favours the second option but definitive proof is awaited (Gekas, *et al* 2005, Ottersbach and Dzierzak 2005).



**Fig 1.2: Sites and types of haematopoiesis during murine fetal life**

The first recognisable blood cells are seen at 7.5 days post-coitus (days post-coitus = E) emerging in the extra-embryonic yolk sac and forming blood islands with production of primitive erythroblasts (top of figure). By E9.5 the first definitive haematopoiesis is emerging (middle of figure). By E10.5 the major site of haematopoiesis has switched to the Aorta-Gonad-Mesonephros (AGM) region of the dorsal aorta with evidence of haematopoietic intra-aortic clusters (HIACS) of cells and expression of haematopoietic genes in sub-aortic patches (SAPs). By E12 production has switched to the fetal liver (bottom of figure). This remains the major haematopoietic organ until after birth when a shift occurs to bone-marrow derived haematopoiesis.

### 1.1.1.2 Primitive and definitive haematopoiesis

In addition to changes in the site of haematopoiesis, the nature of the cells produced also varies during ontogeny. These cells are produced in two overlapping waves termed primitive and definitive haematopoiesis.

*Primitive haematopoiesis:* As discussed above the first wave of blood formation in the yolk sac produces a transient population of primitive erythrocytes that are large and nucleated. The progenitors of these cells are detected between E7.25 and E9 in mouse embryos (day 15-18 in humans). Primitive pro-erythroblasts enter the newly formed embryonic circulation at E8.25 and undergo maturation resulting in an increase in haemoglobin, reduction in size and nuclear condensation. In the mouse they display active globin switching from early  $\beta^H$ -1 expression to later  $\epsilon\gamma$  expression and an  $\alpha$ -globin cluster change from  $\zeta$  to  $\alpha$ -1 and  $\alpha$ -2 expression. In fact all of the globin genes from the alpha ( $\zeta$ ,  $\alpha$ -1 and  $\alpha$ -2) and beta ( $\epsilon\gamma$ ,  $\beta^H$ -1,  $\beta$ -1 and  $\beta$ -2) globin clusters are expressed in primitive erythrocytes (Leder, *et al* 1992, Trimborn, *et al* 1999). This maturation is accompanied by enucleation and these circulating primitive erythrocytes can persist until a few days after birth (Kingsley, *et al* 2006).

It has long been assumed that primitive haematopoiesis is unilineage but evidence is now emerging for other primitive lineages. Primitive macrophage production has been observed to accompany the first wave of primitive haematopoiesis with cells positive for the mouse pan-macrophage marker F4/80 being detectable in early blood islands (Naito, *et al* 1996). Evidence has also emerged for the existence of primitive megakaryopoiesis (Tober, *et al* 2007, Xu, *et al* 2001). This would make developmental sense as the ability to transport nutrients and oxygen to tissues in the very early embryo must depend on red cells, the formation of a vascular system, and a method to maintain its integrity and prevent bleeding. Indeed knock-out mice generated to look at genetic control of very early stages of haematopoiesis usually die in early embryonic life, and death is often accompanied by anaemia and haemorrhage. The latter could be a function of inadequate vasculogenesis or inadequate haemostasis or a combination of the two. Megakaryocytes are first detectable at the same time, and in the same location, as primitive erythroid precursors. Single colonies contain both primitive erythroid and megakaryocytic cells suggesting a common primitive meg-erythroid progenitor (MEP). This has been confirmed using serial dilution experiments to show that a single progenitor cell can give rise to both lineages. Platelets are first detected at E10.5 and early platelets appear larger than those from later stage embryos and adults and are reticulated (Tober, *et al* 2007). These observations are

particularly important when considering the potential target cell for mutations in childhood acute megakaryocytic leukaemia as discussed in section 1.3.2.

*Definitive haematopoiesis:* The second wave of blood production leads to HSCs capable of long-term repopulating ability in adult recipients and has therefore been termed definitive haematopoiesis. These HSC proliferate extensively in fetal life and then sustain blood production throughout life. Definitive haematopoietic activity is first detected at about E10.5-12.5 in mice (day 35-42 in humans) and arises in a number of regions. Definitive erythrocytes differ phenotypically from primitive erythrocytes- they are smaller and usually enucleate before entry into the bloodstream. In mice only the adult globin genes ( $\alpha$ -1,  $\alpha$ -2,  $\beta$ -1 and  $\beta$ -2) are expressed in definitive cells (Leder, *et al* 1992, Trimborn, *et al* 1999). Humans undergo an additional embryonic to fetal ( $\gamma$ -globin) switch before switching to adult  $\beta$  chain around the time of birth and may express fetal haemoglobins even in adult life (Stamatoyannopoulos 2005).

Although often equated with intraembryonic HSCs (i.e. those generated in the AGM region/dorsal aorta discussed in 1.1.1.1), the yolk-sac also contributes to definitive haematopoiesis. Yolk-sac derived HSCs are capable of some haematopoietic reconstitution of adult recipients as shown by recent cell tracking experiments in mice where yolk-sac HSCs pulse-labelled using inducible Cre-recombinase activity (see chapter 4.2.2.2) at E 7.5 were found to comprise 1-10% of all the major peripheral blood cells and erythroid precursors in the bone marrow (BM) in 9-12 month old mice (Samokhvalov, *et al* 2007). Whether yolk-sac derived precursors can contribute to lymphoid reconstitution remains controversial (Matsuoka, *et al* 2001, Yoder, *et al* 1997). Some of this may be explained by the methods used to assess developmental potential of yolk-sac derived precursors – plasticity is a recognised feature of stem cell populations and therefore experimental manipulation such as transgene expression and transplantation may take cells out of their normal context and allow them to display properties not pertinent to their normal role in vivo (Herzog, *et al* 2003). Overall it seems likely that yolk-sac progenitors are capable of producing mature blood cells of at least the erythroid and myeloid lineage in the fetus but have only a limited capacity to contribute to adult haematopoiesis in normal development. In fact, it has recently been postulated that yolk-sac (and possibly the more mature haematopoietic cells associated with HIACs in the AGM region- see 1.1.1.1 above) precursors may be responsible for the majority of fetal blood production. This would allow AGM derived definitive HSCs to devote themselves to self replication- necessary to create a large stem cell pool to sustain haematopoiesis throughout adult life- rather than diverting down differentiation pathways to sustain fetal haematopoiesis (Pina and Enver 2007). This

model is given weight by emerging evidence for distinct properties of LTR-HSCs isolated before or after birth. Fetal HSCs are highly proliferative cells undergoing mainly symmetrical self renewal on division. These cells appear to undergo an intrinsically regulated switch at around 3-4 weeks postnatally that changes them into largely quiescent cells which undergo asymmetric division producing differentiated progeny, in addition to self renewal (Bowie, *et al* 2007).

### 1.1.1.3 Haematopoietic niches

The evidence reviewed in 1.1.1.1 describes the various sites for generation of HSCs. Once generated these HSCs need to settle in permissive sites that allow them to survive, expand and differentiate to form mature blood cells. Although the yolk-sac can support erythroid and megakaryocytic differentiation, there is little evidence for differentiation of mature haematopoietic cells in the AGM region. Instead these cells appear to home to successive organs which act as sites of expansion and differentiation – initially the fetal liver at E10, then the thymus (T cell development) and spleen (predominantly B cell and macrophage development), and finally to the bone marrow. Bone marrow haematopoiesis is first detected at E17.5 in the mouse (week 20 in the human) and by the time of birth is the major site of haematopoiesis. As discussed above the placenta may also represent a site for expansion (and/or generation) of HSCs and again these cells appear to home to fetal liver before differentiating. The molecular basis for homing of stem cells is not fully understood although there is evidence that the chemokine receptor CXCR4, expressed on HSCs, and its ligand CXCL12 (also known as stromal cell derived factor 1 / SDF-1) play important roles in directing homing to BM and localisation of HSCs within supportive microenvironments (Sugiyama, *et al* 2006). The molecular basis for fetal liver homing is presently unknown and SDF-1 knockout mice have normal fetal liver myelopoiesis making the CXCR4-SDF-1 system unlikely to play a role here (Nagasawa, *et al* 1996). Expression of the adhesion molecule integrin  $\alpha$ -4 is essential for entry into the fetal liver but is unlikely to direct homing *per se* (Qian, *et al* 2007).

The decision of HSCs to enter cell cycle, self renew or differentiate is not merely a cell autonomous event, additional regulation is achieved by surrounding cells that interact directly via cell to cell contact and indirectly by secretion of various growth factors and signalling molecules. These surrounding cells and signalling molecules are collectively known as the haematopoietic niche. Most observations about the properties of this niche come from studies of bone marrow haematopoiesis. Evidence for a crucial role of mesenchymal cells - particularly osteoblasts - in controlling haematopoietic cell numbers

has been obtained using knockout mice. These studies have concentrated on bone marrow haematopoiesis and have shown crucial roles for genes such as angiopoietin-1 (expressed by osteoblasts and HSCs and responsible for keeping HSCs in a resting multipotent state), as well as Notch and Wnt signalling pathways (Calvi, *et al* 2003, Cumano and Godin 2007). Overall they have shown that the number of HSCs is limited and tightly linked to the number of osteoblasts. In addition, localisation of HSCs in association with the endosteal surface of bone seems to be mediated via a calcium sensing receptor; indeed endochondrial ossification has been shown to be an essential prerequisite for establishment of normal BM haematopoiesis (Adams, *et al* 2006). Other studies have also shown that cells with an HSC phenotype are also in direct contact with the endothelium of sinusoids although the nature of the interaction between these cell types is unknown (Kiel, *et al* 2005). This is an exciting and rapidly growing field. Identification of signals that allow HSC expansion and survival are particularly important for the transplant field which would be revolutionised by the ability to efficiently culture and expand HSCs in vitro.

Unfortunately, very little is known about the properties of the fetal liver haematopoietic niche. HSCs are known to undergo massive expansion in the fetal liver between E12 and E15 so dissection of the signals that govern this would be particularly interesting. Fetal liver does not contain differentiated osteoblasts although it does contain mesenchymal stromal cells capable of differentiation into osteoblasts, chondrocytes and adipocytes in vitro (Mendes, *et al* 2005). It appears likely that these mesenchymal elements will play important roles in controlling HSC activity along with liver endothelial cells and /or currently un-identified cell types.

Even less is known about the fetal spleen. This organ is capable of supporting haematopoietic differentiation from E13-14. It does not support expansion of HSC numbers but provides a supportive micro-environment for the differentiation of haematopoietic precursors. There is some evidence that fetal spleen stroma may preferentially induce certain mature cell types such as macrophages (Bertrand, *et al* 2006).

Homing of HSCs to the bone marrow occurs from E17.5 onwards in the mouse. After birth the bone marrow rapidly takes over as the major site of haematopoiesis although it should be noted that the liver and spleen are also capable of haematopoiesis in adults at times of haematopoietic stress (e.g. severe haemolytic anaemias such as thalassaemia) or bone marrow stromal dysfunction (e.g. primary myelofibrosis). In addition some paediatric haematological disorders such as Juvenile Myelomonocytic Leukaemia are associated with fetal type haematopoiesis and massive hepatosplenomegaly.

### **1.1.2 The haematopoietic hierarchy**

The discussion above has concentrated on where HSCs are generated during development and how their ability to produce mature progeny may vary during ontogeny. This next section will focus on the nature of the haematopoietic stem cell and the process of differentiation into mature blood cells.

#### **1.1.2.1 The haematopoietic stem cell**

Haematopoietic stem cells sit at the top of the haematopoietic hierarchy and are able to both self-renew and produce all mature blood cell types. They are defined by their ability to stably reconstitute the entire haematopoietic system upon transplantation to appropriately conditioned adult recipients – the classic assay for HSC being long-term reconstituting activity (LTR activity). HSCs have a number of possible developmental options including apoptosis, self renewal, proliferation, quiescence or differentiation. The molecular processes that underlie these developmental fate decisions are discussed below. At a cellular level haematopoiesis is often depicted in a hierarchical fashion with HSCs giving rise to progenitors which then go on to mature along single or multiple pathways (Reya, *et al* 2001) as shown in figure 1.1. Although useful, this model is undoubtedly an oversimplification and may not reflect the true situation in vivo.

#### **1.1.2.2 Progenitor cells**

The progenitor cell compartment consists of all the intermediate progeny between multipotent stem cells and mature differentiated cell types. Progenitor cells are classically described on the basis of their ability to form colonies in vitro. The earliest progenitor cells retain some stem cell characteristics in that they are capable of repopulating an adult organism producing all the different mature blood cell types but cannot maintain long term activity. These cells are called short-term HSCs (ST-HSCs). Next, multipotent progenitors (MPPs) emerge- these are capable of in vitro multi-lineage output but cannot repopulate mice in vivo, probably reflecting their limited proliferative capacity. Classically the earliest true lineage division was thought to occur between the lymphoid and myeloid compartments. In vitro assays identified committed progenitors capable of forming large colonies containing multiple maturing myeloid cell types (mixed colony forming cells – CFC, also known as the common myeloid progenitor - CMP) but which do not contain any lymphoid cells (Akashi, *et al* 2000, Kondo, *et al* 1997). Lymphoid cells are thought to arise from their own common lymphoid progenitor (CLP) (Kondo, *et al* 1997). The CMP goes on to produce cells capable of progressively more restricted outputs such as granulocyte-

macrophage CFC and eosinophil CFCs. Progenitors also become less proliferative, producing progressively smaller colonies, although the overall number of progenitors in each compartment increases due to successive amplification steps.

More recently, progenitor cells and the various HSC populations have been further classified by their expression of cell surface antigens (usually detected by flow cytometry) and at the level of gene expression. This has allowed much more sophisticated dissection of the various intermediate progenitor populations. The model of a complete and irreversible split between myeloid and lymphoid compartments as the first step in lineage commitment has recently been challenged (Adolfsson, *et al* 2005). A population of immunophenotypically defined ST-HSCs (LSK+CD34+Flt3+ cells) possess, at the single cell level, a combined myeloid and lymphoid differentiation potential. However some experimental evidence suggests they have lost the capacity to produce megakaryocytic and erythroid cells (Adolfsson, *et al* 2005). This suggests that the first lineage split is into a meg-erythroid precursor versus a common myeloid-lymphoid precursor. This model has been challenged by other groups (Forsberg, *et al* 2006). However, cell populations which are myeloid-lymphoid primed with multilineage potential, but with little or no meg-erythroid potential, do appear to be detectable by analysis of the expression of key lineage specific transcriptional regulators at the single cell level in both fetal liver and adult bone marrow (Månsson, *et al* 2007). Overall the balance of evidence seems to favour the early emergence of meg-erythroid precursors during the transition from HSC to committed progenitors. In fact, this debate probably illustrates some of the difficulties in using a cellular hierarchy model to explain haematopoiesis. Intermediate populations may not exist as distinct entities with irreversible branch points marking splitting between different lineages. There may instead be a progressive change in the balance of transcription factors within the cell that preferentially encourages it along certain pathways (Hu, *et al* 1997).

### **1.1.3 Lineage specification at the molecular level**

The differentiation of mature lineage-specific cells from multipotent progenitors involves the institution of a specific programme of gene expression. At the molecular level this involves both production of proteins vital for cell function and repression of irrelevant protein production. For example, red blood cells must express haemoglobin to carry oxygen but have no need to express immunoglobulin genes whilst B lymphocytes require high levels of immunoglobulin gene expression but do not need any haemoglobin. Thus two processes are involved in the molecular control of differentiation; gene expression and gene repression. Genes are expressed via transcription of messenger ribonucleic acid

(mRNA) by RNA polymerase II (pol II) and translation of this messenger RNA into protein which may then undergo a number of post-translational modifications. The cell has a number of ways of specifying which proteins it wants to express and those it does not. These include control at the epigenetic level with modifications of chromatin preventing or encouraging entry of the transcriptional machinery. The decision about which individual genes to transcribe into mRNA is generally controlled by transcription factors which bind deoxyribonucleic acid (DNA) upstream of their target genes and direct mRNA synthesis by pol II. Transcription factors often work in combination in the form of a transcription factor complex. Post-transcriptional control also exists, so not every mRNA is destined to be transcribed into protein. In particular, an emerging field is the post-transcriptional control of mRNA stability and translation by a number of regulatory small RNA molecules known as microRNAs (He and Hannon 2004). The discovery of these highly conserved RNA duplexes has had massive implications for the understanding of developmental regulation of gene expression as well as producing powerful tools for the manipulation of gene expression in the laboratory. Downstream of mRNA translation a number of control points exist with post-translational modification of proteins and the ability of the cell to target these proteins to certain areas of the cell or target them for degradation. From this brief overview it can be seen that unravelling the complexities of cellular differentiation is likely to require many different approaches.

### **1.1.3.1 The role of cytokines in differentiation**

Early studies established that cellular differentiation also requires the presence of exogenous cytokines (Metcalf 1970b). Some of these cytokines appear to be lineage specific such as erythropoietin (EPO) for red blood cells and thrombopoietin for platelet production, whilst others appear to act on more primitive progenitors such as granulocyte/macrophage colony stimulating factor (GM-CSF) which acts on earlier myeloid precursors. Two potential models exist for the role of cytokines in cellular differentiation: they could play a permissive role, or an instructive role (Socolovsky, *et al* 1998). In the permissive model the commitment of a progenitor to a particular lineage is a stochastic (random) event, after which differentiation proceeds along a predetermined programme and the role of cytokines is to ensure survival or proliferation of these lineage committed cells. Alternatively, in instructive models activation of cytokine receptors would lead to the institution of a particular programme of differentiation for example EPO signalling would turn on erythroid genes to produce erythroid differentiation. The weight of experimental evidence now favours the permissive model. It has been shown that separating 2 daughter cells after division of a stem cell and then replating them under



identical culture conditions with the same exogenous growth factors leads to very different lineage outputs in the “twin” colonies (Suda, *et al* 1984). This implies that the cellular fates of the daughter cells were not determined by their exogenous culture conditions. Addition of cytokines to culture media does not alter the differentiation potential of stem cells (Metcalf 1991). Also, experiments altering cytokine receptors either by ectopic expression of growth factors in multipotent progenitors or experiments switching the intracellular domain of one growth factor for that of another are unable to redirect lineage choice. Instead the result seems to be to increase proliferation and/or ensure the survival of progenitors and committed cells (Socolovsky, *et al* 1998). Therefore, it appears that cytokine receptor expression is the result rather than the cause of lineage commitment. The production of cytokines would then serve to expand and/or ensure the survival of cells bearing the appropriate receptor allowing the organism to match production of individual lineages to need. Overall, evidence points to transcription factors being the key molecules that determine cellular differentiation – these are discussed below.

### 1.1.3.2 The role of transcription factors in differentiation

Transcription factors are defined by their ability to bind DNA upstream of their target gene at promoters, enhancers or other regulatory sequences. They are classified into a number of groups depending on the nature of their DNA binding domains (Stegmaier, *et al* 2004). The main groups, with representative examples, are listed in Table 1.1.

Class (DNA binding domain)	Examples
<b>Helix-turn-Helix motif:</b>	
<i>Homeodomain</i>	HoxA9
<i>Paired box</i>	Pax-3
<b>Zinc-co-ordinating DNA binding domains:</b>	
<i>4 Cysteine zinc finger of nuclear type</i>	Retinoid X receptors
<i>2 Cysteine 2 Histidine Zinc Finger</i>	Sp1, Ikaros
<i>Diverse Cysteine 4 Zinc Finger</i>	GATA-1.2.3
<b>Basic domains:</b>	
<i>Leucine Zipper</i>	c-Jun
<i>Helix-loop-helix</i>	E2A, MyoD
<b>Beta Scaffold factors with minor groove contacts:</b>	NFkappaB, Runx-1
<b>Miscellaneous</b>	
<i>Pocket domain</i>	P107

**Table 1.1: Classification of transcription factors**

The transcriptional control of haematopoiesis is a huge area of research, this system is regarded as a paradigm for molecular regulation of cellular differentiation. Identification of key transcription factors has come from a number of sources: naturally occurring mutations, the study of induced mutations in invertebrates and lower vertebrates (especially *Drosophila*, zebrafish and *Caenorhabditis elegans*) and the characterisation of chromosomal translocations found in human leukaemias. In fact more than 50% of known transcriptional regulators were discovered as genes involved in leukaemogenesis – haematopoietic cell fate and leukaemia development are closely related (Orkin 2000). Elucidation of the functions of these transcription factors has been heavily dependent on the development of experimental techniques allowing control of gene expression in the mouse (transgenic technology). Two major methods exist for assessing transcription factor function. Firstly the gene can be introduced as a transgene and the effects of ectopic expression assessed in vitro or in vivo. Alternatively a gene knockout can be created where the endogenous copy of the gene is replaced by an inactivated form. These techniques have recently been supplemented by the ability to knock down gene expression using RNA interference (RNAi) technology. Large numbers of transcription factors have been identified and elaboration of the experimental evidence for each factor is beyond the scope of this introduction. Instead examples will be discussed to illustrate key features of the process. The role of GATA-1 in differentiation will be reviewed in section 1.2.

#### ***1.1.3.2.1 Master regulators***

The identification of a number of transcription factors whose expression was limited to particular cell lineages, along with the observation that ectopic expression of some transcription factors could lead to reprogramming of cells from one lineage to another, led to the concept of “master regulators”. These master regulators are proposed to act at cellular hierarchy branch points (nodal points) and determine the decision to proceed along one path or another. This general principle is best illustrated by the key muscle master regulator - myogenic determining factor Myo-D. Ectopic expression of Myo-D is able to induce a complete phenotypic conversion of fibroblasts into myoblasts including expression of key muscle genes and a muscle specific redistribution of organelles (Blau and Baltimore 1991, Davis, *et al* 1987). However Myo-D requires cofactors for this switch as shown by cell fusion experiments. In these experiments ectopic Myo-D is able to stimulate expression of muscle specific genes in fibroblasts but not hepatocytes. Fusion of Myo-D expressing human hepatocytes with native (non-MyoD expressing) mouse fibroblasts results in expression of both human and mouse muscle specific genes. This implies that fibroblasts supply the necessary co-factors needed to institute a muscle pattern of gene expression in hepatocytes (Schafer, *et al* 1990).

In the haematopoietic lineage a number of master regulators have been identified. These genes appear to be capable of lineage specification when ectopically expressed in progenitors or more committed cell types. For example, expression of Maf B in transformed chicken haematopoietic precursors produces monocytic differentiation (Kelly, *et al* 2000). Other master regulators identified in haematopoiesis are listed in table 1.2.

#### ***1.1.3.2.2 Concentration dependent effects***

A number of transcription factors have been shown to have concentration dependent effects. For example, using purified fetal liver haematopoietic progenitors from PU.1 knockout mice it was shown that transfection of these cells with a green fluorescent protein (GFP) tagged PU.1 expressing retroviral vector led to production of macrophage and B cell colonies. A low level of PU.1 expression was associated with a B cell fate and a high level with formation of macrophage colonies (DeKoter and Singh 2000). The mechanisms by which concentration of a transcription factor influences its output are currently unknown but are consistent with the hypothesis that transcription factors form networks with other regulatory proteins and that the overall balance of positive and negative factors may be important in determining a cells fate.

Transcription factor	Expression pattern	Effects of Over-expression	Mouse Knockout Phenotype
GATA-1	Prog, E, Meg, Eos, Mast, Dendritic cells	Increased E, Meg, Eo, Decreased myeloid	Blocked meg and E
GATA-2	Prog, Meg, Mast	Reduced mature E	Decreased Prog
GATA-3	Prog, T cells, Th2	Increased Th2, reduced Th1	No T cells
PU.1	Prog, myeloid, B cells	Increased myeloid	No Myeloid T or B
FOG-1	Prog, E, Meg, Mast	Reduced Eos	Blocked E, no Meg
C/EBP $\alpha$	Myeloid, eos	Increased Eos	No neutrophils & Eos
MafB	Monocyte	Increased monocytes	NK
Runx-1	Haematopoietic cells		No definitive haematopoiesis
Pax-5	B cells	Reduced other lineages	No B cells
Ikaros	Prog, T cells		No lymphoid cells

**Table 1.2: Master Haematopoietic Transcription Factors.** Abbreviations: Prog, progenitors; E, erythroid; Meg, megakaryocyte; Eos, eosinophil; blocked, arrested development; NK, not known - adapted from (Orkin 2000)

#### ***1.1.3.2.3 Negative as well as positive effects on differentiation***

There is evidence that master regulators not only switch on a programme of gene expression but also suppress alternative choices. This is consistent with their proposed role at branch points in the cellular hierarchy – essentially providing a binary switch between two cell fates. As mentioned previously this is undoubtedly an oversimplification, but experimental evidence does exist for concurrent activation and repression of genes as discussed for GATA-1 below in section 1.2.4 (Rodriguez, *et al* 2005).

#### ***1.1.3.2.4 The importance of cellular context and timing of expression***

As shown by the cell fusion experiments with Myo-D described in 1.1.3.2.1 (Schafer, *et al* 1990) a cell must express appropriate co-regulators and/or lack repressors in order to be able to respond to a transcription factor. The timing and site of expression of transcription factors is also crucial. This is particularly pertinent to haematopoiesis which varies in site and nature during ontogeny. Some transcription factors appear to be absolutely required for both primitive and definitive haematopoiesis such as GATA-2 and Scl, whilst others can selectively affect one or the other; Runx-1/AML-1 and c-myb produce severe defects in definitive haematopoiesis whilst leaving primitive haematopoiesis unaffected (Shivdasani and Orkin 1996). In addition, the effect of a transcription factor can differ between adult bone marrow haematopoiesis and fetal liver definitive haematopoiesis. It is important to be aware of these differences when interpreting the results of murine knock-out models as the early effects on blood formation, even in the fetal liver, could arise from a crucial role in yolk-sac derived haematopoiesis or intra-embryonic derived haematopoiesis, failure to dissect out which system is involved could lead to erroneous conclusions about the cause of the phenotype (Pina and Enver 2007).

#### ***1.1.3.2.5 Altering specificity by use of isoforms***

Many key regulators exist as a number of different isoforms. The production of different isoforms is usually achieved by alternative splicing of mRNA incorporating different exons into the final protein product. Transcription factors may utilise different isoforms at different stages of differentiation within a cell type, at different stages of development, or in completely different cell types. This use provides some economy of effort – allowing a constant DNA binding motif to interact with the same set of genes, whilst varying the association domains with other proteins to produce varied outcomes. For example, studies looking at transcriptional regulation of the granulocytic gene myeloperoxidase (MPO) have shown that the transcription factor C-EBP binds to the MPO enhancer in both multipotent progenitors (MPPs) and more committed myeloid cells. However, different isoforms are used in these two populations with predominance of the C-EBP $\alpha$  isoform in MPPs and C-EBP $\beta$  in myeloid cells (Ford, *et al* 1996). It is often impossible to selectively knockout

isoforms using transgenic technology, making analysis of their contributions to cellular differentiation particularly challenging.

### **1.1.3.3 Other methods of control of gene expression**

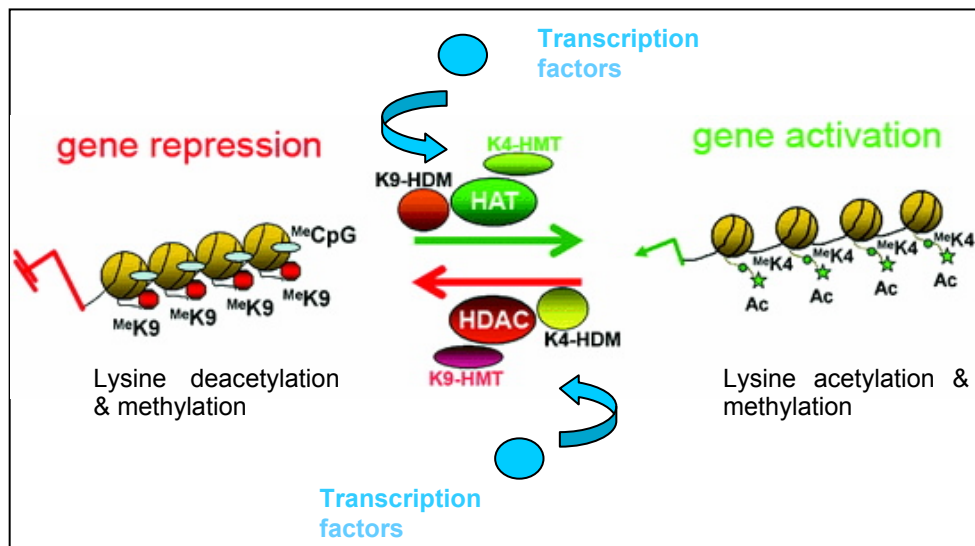
Transcription factors are not the only means by which a cell can regulate gene expression. Detailed discussion of epigenetic, post-transcriptional and post-translational factors is beyond the scope of this thesis but an appreciation of their potential roles is essential when interpreting studies looking at cellular differentiation. They are briefly reviewed below:

#### ***1.1.3.3.1 Epigenetic regulation of gene expression***

The term epigenetic refers to changes in the gene expression profile that arise during development, are heritable, but do not involve alteration of the DNA sequence. The two principal mechanisms of epigenetic control are DNA methylation and histone modifications.

*DNA methylation:* Mammalian DNA can be covalently modified by methylation of cytosine at CpG dinucleotides. This methylation is vital for aspects of developmental biology such as genomic imprinting and X-inactivation (Jaenisch and Bird 2003). More recently reversible DNA methylation has been shown to be important during cellular differentiation; regulation of this methylation appears to be intimately related to chromatin remodelling. Histone methylation, discussed below, is linked to DNA methylation, and may be a mechanism to promote cellular memory of the differentiated state (Bird 2002).

*Histone Modifications:* Chromatin can exist in two opposing states; open or closed. Open active chromatin is accessible for entry of transcriptional machinery and subsequent mRNA synthesis. Closed chromatin is associated with transcriptionally inactive sites. The transition between these two states is governed by modifications of histones, the proteins around which DNA is wound to make up chromatin, as shown in fig 1.3. Methylation of lysine residues by histone methyltransferases (HMTs) also alters transcriptional accessibility (Jaenisch and Bird 2003, Wu and Sun 2006). HATs, HDACs and HMTs can form multi-protein complexes involving transcription factors. For example, Pu.1 and GATA-1 are known to interact and antagonise one another's lineage specific programme of gene expression. This physical interaction leads to recruitment of a repression complex which includes a histone methyltransferase leading to transcriptional silencing of erythroid genes (Stopka, *et al* 2005).



**Fig 1.3. The epigenetic control of gene expression.**

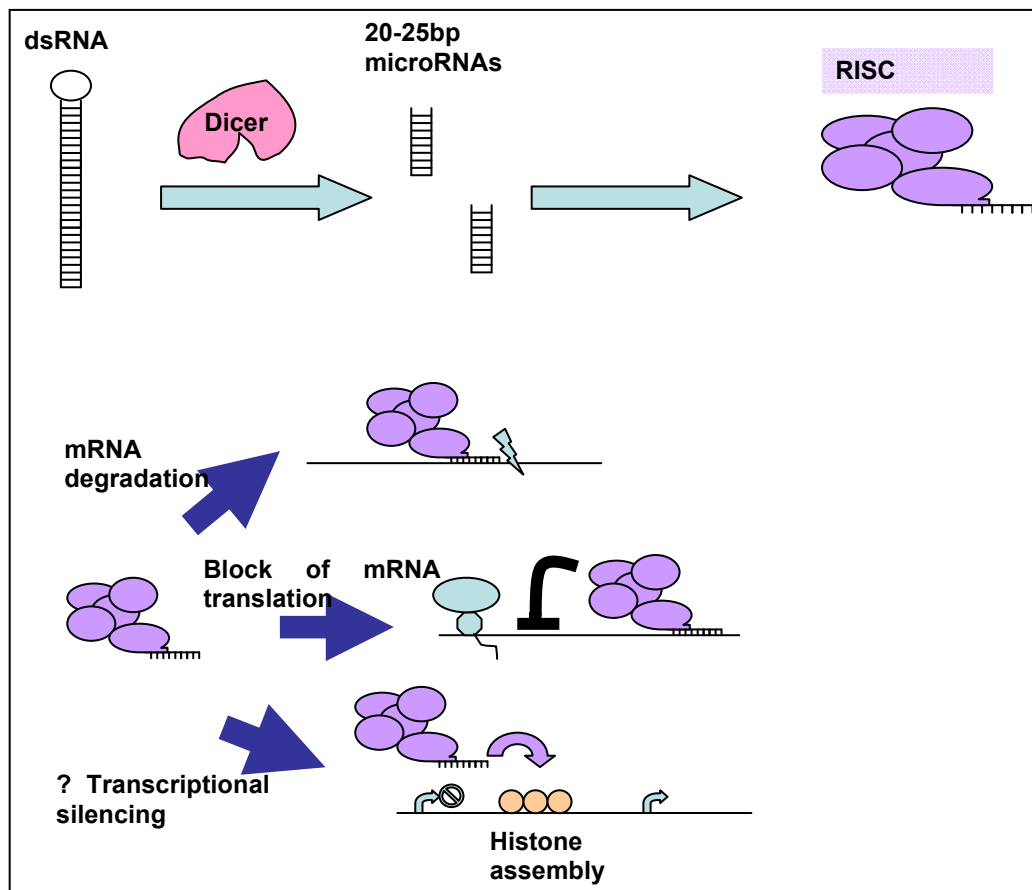
Acetylation of lysine residues by histone acetyltransferases (HATs), marks transcriptionally competent regions and equates with gene activation (left-hand side of figure). Histone deacetylases (HDACs) reverse this activated phenotype, causing gene repression (right-hand side). Methylation of lysine residues by histone methyltransferases (HMTs) also alters transcriptional accessibility. Transcription factors can also interact with HATs, HDACs and HMTs modulating their activity.

#### **1.1.3.3.2 Small non-coding RNAs**

Regulation of gene expression by non-coding regulatory small RNA molecules is a recently discovered but fundamental mechanism by which cells can regulate patterns of gene expression. A number of different classes of small RNAs have been described. The two main classes are small-interfering (si)-RNAs and micro-RNAs. siRNAs take part in an evolutionarily conserved pathway known as RNA interference. This pathway acts as a sequence specific gene silencing process induced by exposure to double stranded RNA, and may have arisen initially as an endogenous defence mechanism against viral infection. Micro-RNAs are related to siRNAs but appear to be encoded in gene-like elements organised in a characteristic inverted repeat in the genome. When transcribed these micro-RNAs give rise to a characteristic stem-looped precursor which is cleaved by an enzyme called dicer into short 21-25bp RNA duplexes (He and Hannon 2004). The proposed mechanisms of gene silencing by micro-RNAs are illustrated in figure 1.4.

Although incompletely understood, miRNAs have been shown to play key roles in haematopoiesis. This is probably best illustrated by the discovery of miR-181 (Chen, *et al* 2004). This micro-RNA was found to be expressed in murine haematopoietic cells and levels of expression were dynamically regulated during lineage commitment. Higher levels were seen in B cells and ectopic expression of miR181 in HSCs led to increased B cell numbers in vitro and in vivo (Chen, *et al* 2004). Later studies have shown crucial roles for

miR-223 in granulopoiesis (Johnnidis, *et al* 2008). The authors also showed that one of the targets of miR-223 is Mef-2c a transcription factor that normally promotes myeloid progenitor proliferation. This underscores the recurrent theme that transcription factors are just one part of a network of gene regulatory factors whose net effect is to repress or activate functional protein production by the cell. Deregulated micro-RNA expression has also been associated with haematopoietic malignancy (Garzon and Croce 2008).



**Fig 1.4 Mechanisms of gene regulation by micro-RNAs**

Transcription of micro-RNAs gives rise to a characteristic stem-loop precursor which is cleaved by an enzyme (*dicer*) into short 20-25bp RNA duplexes. These duplexes interact with a multiprotein complex; RISC (RNA-induced silencing complex). The duplex provides sequence specificity and the associated proteins exert gene silencing via one of 3 mechanisms 1) cleavage of mRNA 2) blockade of translation 3) transcriptional silencing via histone assembly.

#### **1.1.3.3.3 Post-transcriptional control of gene expression**

The detection of mRNA in a cell does not necessarily equate to the production of functional protein. Micro-RNAs are only one of many mechanisms that a cell possesses to regulate the degradation of mRNA or facilitate its translation into protein. Other control mechanisms include processes that regulate turnover of RNA and processes that provide quality control by the degradation of faulty transcripts. Although the basic mechanisms of RNA processing are well described, the degree to which they are dynamically regulated

and gene specific are not well understood (Mata, *et al* 2005). Another level of control is at the level of translation into protein. Precise mechanisms that control gene specific translation rather than global effects on cellular protein production are largely unknown, with the exception of microRNAs discussed above (Gebauer and Hentze 2004). Finally, a cell can control protein activity by a number of post-translational modifications such as phosphorylation, acetylation and sumoylation. These may modify interactions with other proteins or DNA or target the protein for degradation via the ubiquitin-proteasome pathway (Hernandez-Hernandez, *et al* 2006). These are discussed with reference to GATA-1 below. A better understanding of gene-specific post-transcriptional control is likely to provide more insights into mechanisms of cellular differentiation in the future.

## **1.2 GATA-1: a key haematopoietic transcription factor**

The general themes outlined above are well illustrated by a more in depth analysis of the function of a master regulator in haematopoiesis – GATA-1.

### **1.2.1 The GATA family of transcription factors**

The GATA family comprises 6 known transcription factors GATA 1-6. They all share a common mechanism of DNA binding, namely the use of two characteristic C<sub>4</sub> zinc-finger motifs (unique to the family), to bind to a DNA consensus sequence (A/T)GATA(A/G). The DNA binding regions are highly homologous between the members but the rest of the molecule is divergent. The GATA factors fall into two broad subgroups – those involved with haematopoiesis: GATA1, 2 and 3, and those not expressed in haematopoietic tissues: GATA 4, 5, 6. The latter are expressed in several tissues including intestine, heart and lung (Ferreira, *et al* 2005). GATA-1 (also known as NF-E1, ERYF-1 and GF-1) was the founding member of this family and was first identified as a protein binding to the  $\beta$ -globin 3'enhancer- reflecting its fundamental role in erythropoiesis (Wall, *et al* 1988).

### **1.2.2 Cloning and initial characterisation of GATA-1**

GATA-1 was cloned in 1989 using a cDNA expression library constructed from mouse erythroleukaemia (MEL) cells (Tsai, *et al* 1989). Further characterisation showed GATA-1 to be expressed in both the primitive and definitive erythroid lineages, megakaryocytic, eosinophil and mast cell lineages (Ferreira, *et al* 2005). More recently a role in the dendritic cell lineage has been described (Gutierrez, *et al* 2007). It was also shown to be



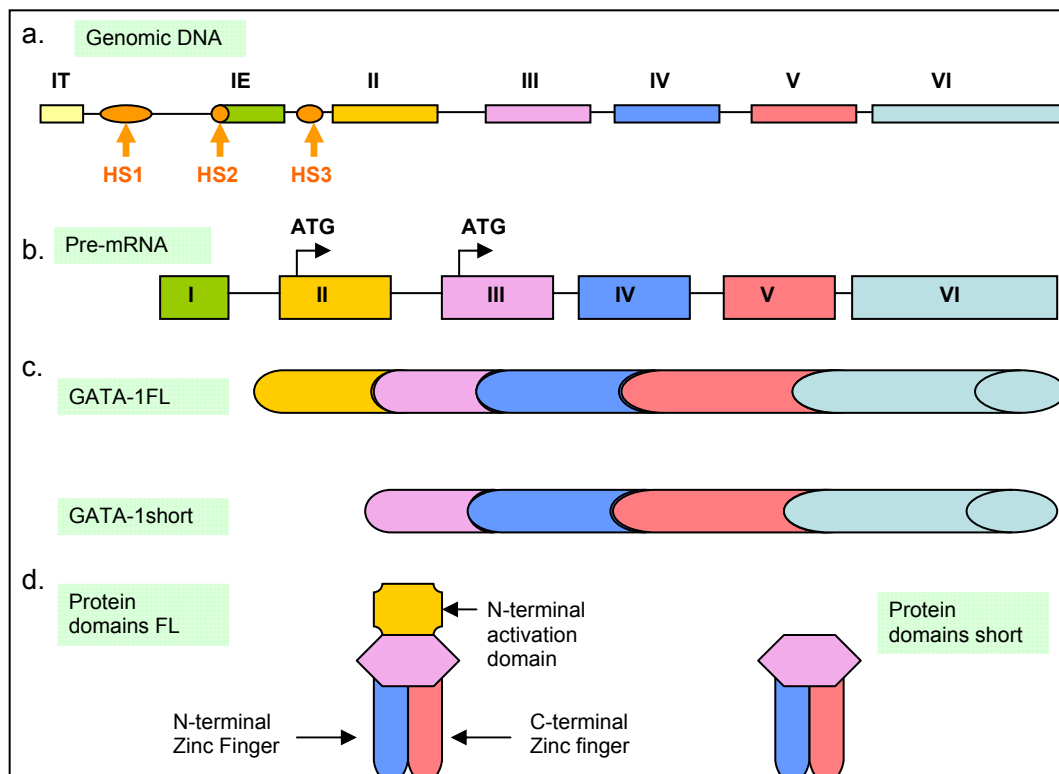
abundantly transcribed in the Sertoli cells of the testis although this utilised a different promoter associated with a testes-specific non-transcribed first exon (known as IT to contrast with the haematopoietic first exon IE) (Ito, *et al* 1993). GATA-1 is highly conserved between mammalian and avian species including humans, mice and chickens and a GATA-1 equivalent has also been cloned in Zebrafish (Lyons, *et al* 2002). Both murine and human GATA-1 genes are located on the X chromosome (Zon, *et al* 1990).

*Structure:* GATA-1 has at least two isoforms – the common full length isoform known as GATA-1FL and a shortened version of the protein GATA-1s. GATA-1FL encodes a 413 amino acid protein with at least 3 functional domains: an N terminal activation (or repression) domain, an N-terminal zinc finger and a C-terminal zinc finger. GATA-1s lacks the N terminal activation domain. The intron-exon structure and protein domains are illustrated in Fig 1.5. The N-terminal activation domain is important for interaction with other proteins and is discussed in detail in section 1.2.8 below. The C-terminal Zinc finger is responsible for recognition of, and binding to, the GATA-1 consensus sequence on DNA. The N-terminal zinc finger does not bind DNA directly but contributes to the stability and specificity of this binding. In addition the N-terminal zinc finger is responsible for interaction with important GATA-1 cofactors including Friend of GATA-1 (FOG-1). Some of these factors associate exclusively with the N-finger whilst others, such as erythroid krüppel-like factor (EKLF), bind both the N- and C-fingers. There is also evidence that GATA-1 can homodimerise (Crossley, *et al* 1995), which may be important in the regulation of promoters with multiple GATA binding sites.

*Transcriptional Control:* GATA-1 activity is highly regulated and is controlled at the transcriptional level by different mechanisms in different cell types. DNase I hypersensitivity analysis of the GATA-1 locus in erythroid cells reveals 3 main hypersensitive sites termed HS1, 2 and 3. HS1 lies between 3.9 and 2.6 kilobases (kb) upstream of the erythroid first exon (IE). HS2 corresponds to the putative GATA1 haematopoietic promoter in the region of IE. HS3 is located in the first intron. These are also shown in Fig 1.5.

The various GATA regulatory elements contain duplicated GATA binding sites and cotransfection experiments have shown that GATA-1 or other family members are able to regulate GATA-1 transcription (Hannon, *et al* 1991, Nicolis, *et al* 1991, Tsai, *et al* 1991) suggesting auto-regulation as at least part of the mechanism of transcriptional control. Studies looking at the individual function of these regulatory regions were carried out by linking cloned fragments of the three control sites to a LacZ reporter gene and expressing

these in transgenic mice. The proximal HS1 element was sufficient for faithful recapitulation of GATA-1 expression patterns in primitive haematopoiesis and revealed very early GATA-1 transcription in E7.0 embryos. However this element was insufficient for expression in definitive haematopoietic cells. Expression of the reporter gene in these cells required both the HS3 intronic element and the HS1 region (Onodera, *et al* 1997). Interestingly the HS3 intronic region does not behave like a classical enhancer or promoter and is unable to transactivate reporter genes when used alone. The mechanism by which this element activates transcription is still unknown (Shimizu and Yamamoto 2005).



**Fig 1.5 GATA-1 Structure**

**a) genomic DNA locus showing location of DNase hypersensitive sites (not to scale) b) primary mRNA transcript showing location of two alternative start codons. Coloured blocks represent exons and black lines represent introns c) alternative splicing or translation produces two polypeptide chains corresponding to the two isoforms GATA-1FL and GATA-1s d) protein structure showing the two zinc fingers and the N-terminal domain.**

Further studies have shown that lineage specific GATA-1 expression requires different regions of HS1, with megakaryocytic expression requiring the whole 317bp sequence whilst erythroid expression can be recapitulated using only the 5' 62bp (Nishimura, *et al* 2000, Vyas, *et al* 1999b). Eosinophil expression requirements vary yet again being independent of the HS1 region (Guyot, *et al* 2004) but requiring a double GATA motif in HS2 (Yu, *et al* 2002a). There are likely to be several other cis-regulatory elements not yet fully characterised and it appears that, despite significant areas of homology between

murine and human GATA-1 in terms of function and regulation, the chromatin structures between the mouse and human loci vary significantly (Valverde-Garduno, *et al* 2004).

*Post-transcriptional and post-translational control:* A number of post-translational modifications to the GATA-1 protein have been reported. These include acetylation, phosphorylation and Sumoylation. GATA-1 acetylation has been described in a wide variety of species with variable effects on function although overall it appears that acetylation mutants are associated with a reduced ability of GATA-1 to rescue GATA-1 deficient phenotypes in model systems (Boyes, *et al* 1998, Ferreira, *et al* 2005). It has been proposed that the combination of acetylation and phosphorylation of GATA-1 targets the protein for degradation via the ubiquitin-proteasome pathway (Hernandez-Hernandez, *et al* 2006). GATA-1 phosphorylation patterns seem to vary during induction of haematopoietic differentiation in cell lines (Crossley and Orkin 1994). In addition, phosphorylation of GATA-1 seems to be dynamically regulated in response to cytokine-induced signalling in haematopoietic progenitors (Kadri, *et al* 2005, Towatari, *et al* 2004). However, the generation of transgenic mice with serine to alanine mutations at sites of phosphorylation revealed that prevention of phosphorylation is still compatible with normal haematopoietic development as measured by peripheral blood counts. More in depth analysis of these mice revealed mild defects in erythroid progenitor cell numbers in adult bone marrow but no functional impairment even in response to acute erythropoietic stress (Rooke and Orkin 2006). Finally, Sumoylation involves the ligation of a small ubiquitin-like peptide (SUMO) to a lysine residue on a target protein. Unlike ubiquitination this does not target a protein for degradation but rather seems to affect localisation, stability and interactions with other proteins. Sumoylation has been shown to be important in regulating the activity of several key cellular proteins including p53 and c-Jun. GATA-1 is sumoylated in vivo but non-sumoylatable mutants have no detectable loss of activity in reporter gene assays. This suggests this modification does not play a major part in affecting transcriptional function, although a more subtle role in regulation is possible (Collavin, *et al* 2004).

Initial insights into GATA-1 function relied heavily on assessing the effects of mutation or ablation of GATA-1 in knockout mice and ES cells as described below. The crucial role of GATA-1 in megakaryocytic and erythroid development has been further reinforced by descriptions of human diseases due to inherited or acquired mutations in GATA-1 including thrombocytopenia, anaemia and acute megakaryoblastic leukaemia.

### **1.2.3 GATA-1 plays a key role in erythropoiesis**

GATA-1 DNA binding sites are found upstream of virtually all erythroid expressed genes leading to early predictions that it may play a key role in erythropoiesis (Evans, *et al* 1988). Binding of GATA-1 to these sites led to strong transcriptional activation using *in vitro* reporter assays (Martin and Orkin 1990). To further investigate the role of GATA-1 in haematopoiesis, GATA-1 knock out embryonic stem (ES) cells were generated and tested for their ability to contribute to different tissues in a chimaeric mouse (Pevny, *et al* 1991). The mutant ES cells contributed to all non-haematopoietic tissues tested and produced some white blood cells but were completely unable to contribute to mature erythroid cells. This study established that GATA-1 had a direct role in haematopoiesis and that other proteins were unable to compensate for GATA-1 deficiency. It did not, however, establish the stage at which erythropoiesis was blocked or the effect of GATA-1 ablation on presumptive target genes.

#### **1.2.3.1 Effects of GATA-1 deficiency in vitro**

To address this, the knockout cells were utilised in *in vitro* ES cell differentiation studies. These showed that GATA-1 null ES cells fail to generate primitive erythroid precursors in cultures of embryoid bodies (Simon, *et al* 1992) or secondary replating experiments looking for emergence of typical tight haemoglobinised primitive colonies in the presence of erythropoietin (Keller, *et al* 1993, Weiss, *et al* 1994). Generation of definitive cells appeared initially normal but these cells were unable to mature, becoming blocked at the pro-erythroblast stage of differentiation followed by cell death (Weiss, *et al* 1994). Surprisingly the arrested GATA-1 null definitive pre-erythroblasts actually expressed putative GATA-1 target genes at normal levels despite their complete lack of GATA-1 expression. In particular, they expressed all the embryonic and adult globin transcripts. Although not interpreted as such at the time, this may actually have reflected the presence of primitive erythroid precursors rather than definitive as the latter are not thought to express embryonic globins (Trimborn, *et al* 1999). Alternatively it could reflect more promiscuous globin gene expression in definitive cells in the absence of GATA-1. The related family member GATA-2 was massively up-regulated in these cells suggesting that GATA-2 is able to compensate partially for GATA-1 in early, but not late, erythroid progenitors (Weiss, *et al* 1994). In the absence of GATA-1, proerythroblasts die by apoptosis (Weiss and Orkin 1995). The mechanism behind this is not fully understood but phosphorylated GATA-1 can upregulate expression of the cellular survival gene E4bp4

and result in increased levels of the anti-apoptotic gene Bcl-X<sub>L</sub> (Yu, *et al* 2005). This suggests GATA-1 plays roles in cellular survival pathways as well as maturation.

### **1.2.3.2 GATA-1 knockout embryos die due to severe anaemia**

5 years after the first description of generation of GATA-1 knockout ES cells, and their use in chimaeric embryos, the results of germline transmission were reported (Fujiwara, *et al* 1996). Male GATA-1 null embryos die between E10.5-E11.5 due to severe anaemia. Prior to this (E9.5) they showed extreme pallor but had detectable erythroid cells, arrested at an early pro-erythroblastic-like stage. Female heterozygotes are anaemic at birth but recover during the neonatal period, presumably because of selection for erythroid progenitors expressing the normal GATA-1 allele. These studies confirmed the essential role of GATA-1 in erythropoiesis but suggested that, in contrast to the *in vitro* experiments, both primitive and definitive erythropoiesis were similarly affected by a cellular maturation block.

### **1.2.3.3 Knockdown variants allow survival of embryos**

The existence of knockout ES cells facilitated a number of subsequent experiments looking at the ability of various mutants to rescue erythropoiesis *in vitro* and *in vivo*. Survival of GATA-1 deficient embryos long enough to assess phenotypes has been achieved by generation of two knockdown variants. “GATA-1 low” mice express about 20% of wild-type GATA-1 levels. Although the majority of these mice die at E13.5-14.5 due to anaemia, a small number survive to adulthood. These mice, although anaemic at birth, recover and have a normal lifespan (McDevitt, *et al* 1997b). “GATA1.05” mice express lower levels – about 5% of wild-type- and, as expected, show a more severe phenotype dying between E11.5 and E12.5 with a huge accumulation of undifferentiated erythroid precursors in the yolk sac (Takahashi, *et al* 1997). Interestingly, longer follow up of these mice reveals that they have an increased risk of two types of leukaemia – an early onset erythroblastic leukaemia and a later onset B cell leukaemia. These leukaemias can be prevented by transgenic restoration of GATA-1 expression (Shimizu, *et al* 2004).

One key study using these knockdown mice established the *in vivo* requirements for the different domains of GATA-1 during primitive and definitive haematopoiesis (Shimizu, *et al* 2001). Transgenic mice carrying mutations in the C-terminal zinc finger ( $\Delta$ CF), the N-terminal finger ( $\Delta$ NF) or the N-terminal trans-activation domain ( $\Delta$ NT) were crossed with GATA1.05 mice. The N and C zinc finger mutants failed to rescue the male hemizygous GATA1.05 pups from their expected embryonic lethality. The  $\Delta$ NT mutant (which

corresponds to a naturally occurring isoform of GATA-1; GATA-1s discussed in more detail below) was able to rescue both primitive and definitive erythropoiesis but its ability to do this was strongly dependent on level of expression. Levels equivalent to endogenous GATA-1 production were able to prevent early embryonic lethality up to E15.5 but high levels of expression were needed to allow survival beyond this. This suggests that GATA-1s can support early primitive erythropoiesis in the yolk sac but not definitive erythropoiesis. There appeared to be some functional redundancy between the N-terminal zinc finger and the N-terminal activation domain in that either was able to compensate for the other in primitive erythropoiesis but a compound mutant ( $\Delta$ NTNF) was unable to do so. In addition, the expression of haem biosynthesis genes was normal in primitive erythrocytes from the  $\Delta$ NF mutant but severely impaired in definitive erythrocytes from these same mice (Shimizu, *et al* 2001). Overall these results suggest that primitive erythropoiesis can proceed despite lack of some, but not all, co-factor binding sites (which are known to be present in the N-terminal finger and the N-trans-activation domain) but that definitive erythropoiesis requires more complex interactions.

#### **1.2.4 Mechanisms of GATA-1 activity**

These early descriptions of the phenotypic effects of GATA-1 deficiency were followed by investigation of the underlying mechanisms by which GATA promotes erythroid development. An overview of the effects of GATA-1 on gene expression was achieved using GATA-1 rescue of a GATA-1 null ES cell line (G1E) (Weiss, *et al* 1997), followed by RNA extraction at various time points and analysis using microarray technology (Welch, *et al* 2004). This revealed that GATA-1 expression affected transcript levels of a very large number of genes (far more than the previously known targets of GATA-1) some of which showed rapid change after GATA-1 rescue whilst others showed a lag period presumably reflecting the need for synthesis and assembly of cofactors, or involvement of intermediate genes. In addition, GATA-1 does not merely activate gene expression - many transcripts were down-regulated by GATA-1 expression. Later, a biotinylation-tagging proteomic approach identified protein partners of GATA-1 in erythroid cells using mass spectrometry. Experiments looking at co-purification of proteins in this system showed that GATA-1 was able to form at least 5 distinct activating and repressing complexes in the cell (Rodriguez, *et al* 2005). One of the key repressive complexes consisted of GATA-1, its cofactor FOG-1 and the MeCP1 complex. The latter complex is a methyl-DNA binding protein, histone deacetylase complex important in dynamic chromatin remodelling.

### 1.2.4.1 GATA-1 Cofactors

Discussion of all the known partners of GATA-1 is beyond the scope of this introduction. Instead a few key interactions will be discussed that illustrate important points

*FOG-1*: FOG-1 is a nine zinc finger protein which binds to the N-terminal Zinc finger of GATA-1. Although it has no direct DNA binding activity it seems to be a key member of complexes involved both activation and repression of target genes (Rodriguez, *et al* 2005, Tsang, *et al* 1997). The phenotype of FOG-1 knockout mice closely resembles that of GATA-1 knockout mice (Tsang, *et al* 1998). In contrast to GATA-1 however, FOG-1 deficiency results in complete ablation of the megakaryocytic lineage suggesting additional roles for this factor in this lineage. Proof that FOG-1/GATA-1 association was essential for erythroid differentiation came from the analysis of GATA-1 mutants unable to bind FOG-1 (Crispino, *et al* 1999). Naturally occurring human mutations that disrupt GATA-1 FOG-1 interactions produce a congenital dyserythropoietic anaemia associated with severe thrombocytopenia (Nichols, *et al* 2000). Since this first report, several mutations in the GATA-1 N-terminal zinc finger have been described (Freson, *et al* 2001, Freson, *et al* 2002, Mehaffey, *et al* 2001, Yu, *et al* 2002b). All these affected families have thrombocytopenia but the degree of anaemia reflects the degree of disruption of FOG-1 binding.

*PU.1*: PU.1 is a member of the Ets family of transcription factors and is a master regulatory gene for the lymphoid and granulocytic lineage (Scott, *et al* 1994). Therefore GATA-1 and PU.1 should show reciprocal activation patterns in the two lineages and this indeed appears to be the case. Direct physical interaction between GATA-1 and PU.1 leads to functional antagonism (Nerlov, *et al* 2000, Rekhtman, *et al* 1999). There is evidence that the stoichiometry of these transcription factors determines cellular fate (Rekhtman, *et al* 1999). This suggests that these factors could play a part in lineage specification as well as maturation with small fluctuations in the relative levels of these factors leading to reciprocal inhibition, which in combination with positive autoregulation could lead to an eventual dominance of one over the other (Bottardi, *et al* 2007).

*Krüppel-like factors*: GATA-1 forms a synergistic physical interaction with the related factors Sp1 and erythroid Krüppel-like factor (EKLF). The former appears essential for early embryonic development, and the latter is essential for definitive erythropoiesis. These Krüppel-like factors recognise CACC motifs that often exist in close proximity to GATA-1 binding sites. It has been proposed that EKLF/GATA-1 co-operation is essential for

bringing distant regulatory elements together – perhaps best studied in the case of the  $\beta$ -globin locus which forms an active chromatin hub, comprising multiple regulatory regions, with extensive looping of intervening chromatin (Drissen, *et al* 2004, Letting, *et al* 2003). Again this example illustrates that the effects of transcription factor interaction can be mediated at levels other than direct interaction with RNA pol II.

*Runx-1 (AML-1)*: Runx-1 plays a key role in the emergence of HSCs from the dorsal aorta during early intraembryonic haematopoiesis (Cai, *et al* 2000) with a complete failure of definitive haematopoiesis seen in Runx-1 knockout mice (Okuda, *et al* 1996). In addition, conditional knockouts have shown an important role for this gene in adult haematopoiesis (Ichikawa, *et al* 2004) and megakaryocyte maturation (Song, *et al* 1999). Runx-1 and GATA-1 show functional co-operation and physical interaction in mice and *Drosophila* during haematopoietic lineage commitment (Elagib, *et al* 2003, Waltzer, *et al* 2003). In primary human haematopoietic progenitor culture, Runx-1 is upregulated immediately prior to megakaryocyte differentiation and is downregulated prior to erythroid differentiation (Elagib, *et al* 2003). Using cell lines induced to undergo megakaryocyte differentiation, the same group showed physical interactions between Runx-1 and GATA-1 by co-immunoprecipitation. These interactions were dependent on the N and C terminus of GATA-1 but not the zinc finger domains (Elagib, *et al* 2003). These physical interactions also had functional significance as measured by synergy in activating a megakaryocyte promoter driven reporter gene. Recent evidence suggests that GATA-1 may target Runx-1 for phosphorylation and thereby change its activity from a repressor to an activator of transcription (Elagib and Goldfarb 2007). This illustrates yet another potential mechanism for cross-talk between opposing or synergistic transcription factors.

*Gfi-1b*: Gfi-1b is a transcriptional repressor involved in megakaryocyte and erythroid differentiation. Analysis of GATA-1 interacting proteins in erythroid cells showed that Gfi-1b/GATA-1 complexes bind upstream of the myc and myb genes thought to be important in the proliferation arrest that is observed with terminal erythroid differentiation (Rodriguez, *et al* 2005). These GATA-1/Gfi-1b complexes did not bind other known GATA-1 cofactors such as FOG-1. In addition Gfi-1b was not found binding to other putative GATA-1 targets. Hence, this may represent a distinct mechanism by which GATA-1 could negatively regulate cell cycle genes.

*miR 144 and miR 451*: Reports of modulation of microRNA levels during haematopoietic development are beginning to emerge (Zhan, *et al* 2007). A recent microarray study showed significant upregulation in 11 out of 292 microRNAs following GATA-1 rescue of



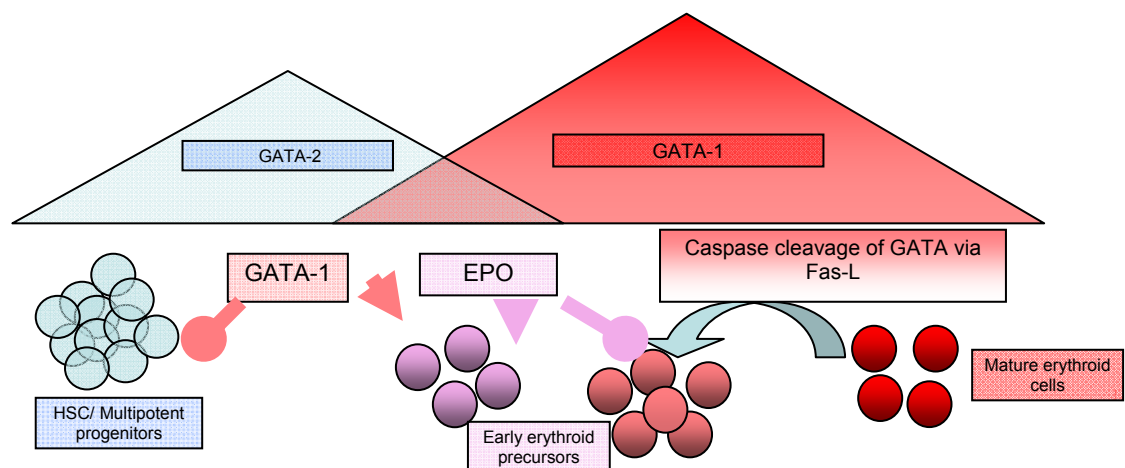
the G1E knockout cell line (Dore, *et al* 2008). The two most strongly upregulated were miR 144 and miR 451. GATA-1 binds to upstream regulatory units and directly controls the expression of these microRNAs. They also showed that zebrafish embryos depleted of miR 451 had abnormal erythroid maturation (Dore, *et al* 2008). The microRNAs had multiple targets but analysis of these showed significant enrichment for nuclear proteins. This lends support to a theory that terminal maturation may be promoted by microRNAs by clearing the cells of transcription factors needed for earlier lineage specification but no longer needed in mature daughter cells. This had previously been suggested by studies in plants where 34/49 identified targets of microRNAs were transcription factors despite these representing only 6% of the coding genes in these species (Rhoades, *et al* 2002).

#### 1.2.4.2 Regulators of GATA-1 expression

*GATA-2*: The second member of the GATA family also plays a pivotal role in haematopoiesis. In contrast to GATA-1, GATA-2 appears to act primarily at the level of the HSC and is important for production and expansion of HSCs in the AGM region and for proliferation of HSCs in adult bone marrow (Ling, *et al* 2004). GATA-2 downregulation appears to be required for haematopoietic differentiation (Persons, *et al* 1999). The observation of up to 50 fold up-regulation of GATA-2 in GATA-1 knockouts first pointed to a possible reciprocal relationship between the two factors (Weiss, *et al* 1994). It appears that the combination of GATA-2 downregulation and GATA-1 upregulation is important for normal erythroid differentiation. GATA-2 binds upstream of its own promoter, recruits a co-factor with histone modification activity (CBP) and induces an open chromatin configuration. GATA-1 has been shown to displace this GATA-2/CBP complex from the GATA-2 promoter, leading to formation of inactive chromatin and loss of GATA-2 expression (Grass, *et al* 2003). GATA-1 also displaces GATA-2 from the  $\alpha$ -globin complex, with the opposite effect of activation of gene expression (Anguita, *et al* 2004).

*GATA-1 (autoregulation)*: GATA-1 expression is dynamically regulated and GATA-1 appears to influence its own expression at transcriptional and post-transcriptional levels. Binding to its own upstream regulatory elements leads to a positive feedback loop whereby GATA-1 levels are rapidly upregulated once a critical level of expression has been achieved. The interaction with GATA-2 described above ensures that this upregulation is co-incident with GATA-2 downregulation. However, GATA-1 levels appear to decline with terminal differentiation (Suzuki, *et al* 2003). In fact, over-expression of GATA-1 in erythroid cells leads to inhibition of differentiation in some circumstances (Whyatt, *et al*

2000). This has been shown to be due to an intrinsic red cell defect but it is not cell-autonomous – in mouse chimaeras the presence of wild-type erythroid cells can rescue overexpressing cells and promote their terminal differentiation (Gutierrez, *et al* 2004, Whyatt, *et al* 2000). Furthermore, it has been shown that activation of death receptors is linked to GATA-1 degradation via caspase mediated cleavage (Ferreira, *et al* 2005). The Fas death receptor is expressed on all erythroid cells but the ligand (Fas-L) is only expressed on mature cells. One possible model is that expression of Fas-L on terminally mature cells feeds back to immature precursors, down regulates GATA-1 levels and encourages them to differentiate rather than proliferate and thereby matches blood cell production to demand. Since this interaction involves homotypic signalling it would explain the rescue of overexpressing cells in the presence of normal mature erythroid precursors. This model has been further complicated, however, by a second observation that GATA-1 can be protected from cleavage by the presence of a nuclear chaperone; Heat Shock protein 70 (HSP70) (Ribeil, *et al* 2007). These authors showed that the presence of EPO caused nuclear co-localisation of HSP70 and GATA-1 and prevention of caspase mediated cleavage. This interaction appeared to be crucial for the prevention of apoptosis in proerythroblasts. A number of overlapping mechanisms appear to operate at different progenitor levels to ensure a matching of mature blood cell output with demand. An overview is shown in Fig 1.6.



**Fig 1.6 GATA-1 autoregulation and cross regulation with GATA-2 during commitment to the erythroid lineage.**

During early haematopoiesis GATA-2 predominates, GATA-2 down regulation and GATA-1 up-regulation are required for erythropoiesis. GATA-1 inhibits GATA-2 and also upregulates its own expression leading to a gradual switch from GATA-2 to GATA-1 production. GATA-1 production peaks in intermediate erythroid precursors before being down-regulated to allow terminal differentiation. This down-regulation is facilitated by caspase mediated cleavage of GATA-1 by activation of Fas death receptors (expressed on all haematopoietic precursors) by Fas ligand (Fas-L) which is only present on terminally differentiated cells. Early haematopoietic precursors are protected from this GATA-1 cleavage via an erythropoietin (EPO) mediated recruitment of nuclear chaperone proteins.

### **1.2.5 The role of GATA-1 in megakaryocyte development**

Megakaryocyte maturation involves a number of specialised processes. As megakaryocytes mature they undergo endoreduplication – the doubling of DNA content in S-phase without subsequent cell division - leading to sequential increases in ploidy with each cell cycle- accompanied by expansion of their cytoplasmic mass. This leads to large cells (up to 100 microns in diameter) with high chromosome numbers (up to 128 N), hence their name – *mega-karyocyte*. Maturation is accompanied by amplification of platelet organelles and demarcation membranes – the precursors of platelet membranes. The cells then form pro-platelet projections which give rise to circulating platelets (Deutsch and Tomer 2006).

#### **1.2.5.1 Generation of transgenic mice with selective loss of GATA-1 expression in the megakaryocytic lineage**

As discussed above megakaryocytes and erythroid cells appear to originate from a common meg-erythroid progenitor (MEP) cell. Initial indications that GATA-1 has crucial roles in the differentiation of both of these lineages came from overexpression studies. Expression of GATA-1 in retrovirally transformed chicken myeloblasts induces differentiation into thromboblats (the avian equivalent of megakaryocytes) as well as erythroid and eosinophil lineages (Kulesa, *et al* 1995). In murine systems, overexpression of GATA-1 in the transformed myeloid precursor line 416B leads to production of megakaryocytes (Visvader, *et al* 1992). In addition the GATA-1 knockout chimaeric mice have large numbers of immature megakaryocyte precursors in their fetal livers (Pevny, *et al* 1995). Since GATA-1 knockout mice die early from lethal anaemia, it is not possible to use this system to examine the requirement for GATA-1 in megakaryocytes. Instead two mutant lines of mice have been generated that exploit the known differences in requirements for upstream regulatory elements for GATA-1 expression in the two lineages (Shivdasani, *et al* 1997). This was achieved by replacement of an 8kb segment upstream of the erythroid first exon of GATA-1, encompassing the testis first exon and the HS1 region, with a neomycin resistance gene flanked by LoxP sites (see chapter 4.2.2.2 for a description of this technology). When transmitted through the germline this mutant strain - called neo $\Delta$ HS - produced a moderate fetal anaemia and reduced platelet levels in hemizygous males. Removal of the neomycin selection cassette by Cre-recombinase expression - to produce a second strain -  $\Delta$ Neo $\Delta$ HS -led to restoration of nearly normal red blood cell production but left the platelet defect unchanged (Shivdasani, *et al* 1997). GATA-1 mRNA is undetectable in megakaryocytes from neo $\Delta$ HS mice but easily detectable in erythroid cells. Low levels of GATA-1 mRNA are detectable in  $\Delta$ Neo $\Delta$ HS

megakaryocytes but the phenotype of the megakaryocytes are the same in both strains suggesting this level is too low to sustain normal megakaryopoiesis. The platelet counts are reduced to about 15% of normal levels in these mice although megakaryocyte numbers are greatly elevated. This increase in numbers is independent of platelet count – heterozygous females also show this defect despite having normal peripheral platelet counts – this rules out a secondary response to thrombocytopenia as the cause of the megakaryocytic hyperplasia. Overall these findings suggest that GATA-1 is a negative regulator of cell proliferation in early megakaryocyte precursors. In addition, electron microscopy confirmed that megakaryocyte precursors are structurally immature with large segmented nuclei, lack of cytoplasmic granules, underdeveloped demarcation membranes and excess rough endoplasmic reticulum (rRER). Proplatelet formation appears blocked and reduced polyploidisation is seen (Vyas, *et al* 1999a). Platelets are similarly abnormal being spherical rather than discoid, large, with few  $\alpha$ -granules and an excess of rRER. These results suggest a second role in cellular maturation. This mirrors the role of GATA-1 in erythroid cells with dual effects on cellular proliferation and terminal maturation.

#### **1.2.5.2 Gene expression mirrors that seen in the erythroid lineage**

Gene expression profiling of  $\Delta$ Neo $\Delta$ HS megakaryocytes versus wild type megakaryocytes confirmed down regulation of some GATA-1 target genes such as NF-E2 (p45 subunit) and GPIb $\alpha$  in GATA-1 deficient cells, but others such as GpIIb were only modestly affected (Muntean and Crispino 2005). This may have been because of an upregulation of GATA-2 (5.5 fold) and Ets-1 (6 fold) – the promoters of many megakaryocytic genes (including GpIIb) have multiple GATA and Ets binding sites. Therefore, like erythroid development, GATA-2 and other transcription factors may be able to partially compensate for GATA-1 loss especially in the early stages of differentiation but are incapable of producing fully mature cellular progeny.

#### **1.2.5.3 Rescue of proliferation and differentiation defects can be uncoupled and mapped to different domains of GATA-1**

Later experiments assessed the ability of various GATA-1 mutant proteins to rescue the megakaryocytic phenotype in  $\Delta$ Neo $\Delta$ HS primary fetal liver progenitor cells in vitro (Kuhl, *et al* 2005). These showed that reversal of the proliferation defect and differentiation defect can be uncoupled. Specifically the FOG-1 binding domain and amino-acid residues 84-110 in the N-terminus are both needed to restore differentiation, whilst residues 54-110 are needed to prevent abnormal growth (Kuhl, *et al* 2005). Of note the naturally occurring isoform GATA-1s lacks residues 1-84 but has normal expression of residues 84 onwards.

This would imply that this variant should have retained differentiation capacity but may not be able to control megakaryocyte growth. This prediction has been strengthened by another set of rescue experiments using  $\Delta$ Neo $\Delta$ HS fetal liver cells cultured in vitro (Muntean and Crispino 2005).  $\Delta$ Neo $\Delta$ HS fetal liver cells were retrovirally transfected with GFP tagged cDNAs encoding wild type GATA-1, GATA-1s (no N-terminal) or a GATA-1 mutant unable to interact with FOG-1. Wild-type GATA-1 was able to restore normal proliferation and maturation of the cultured megakaryocytes. GATA-1s expression led to generation of proplatelets, suggesting effective maturation, but did not correct the hyperproliferative phenotype. The FOG-1 mutant led to reduced proliferation well below that of wild type and a block in differentiation. This result is in contrast to those discussed above where the FOG-1 mutant appeared to have normal growth characteristics. Both these sets of experiments have to be interpreted with caution as substantial overexpression of the exogenous constructs may have influenced the observed phenotypes.

## **1.2.6 Role in other lineages**

### **1.2.6.1 Eosinophils**

The role of GATA-1 in eosinophil development was first suggested by the GATA-1 overexpression studies in avian haematopoietic progenitors, already discussed in relation to erythroid and megakaryocytic differentiation (Kulesa, *et al* 1995). Interestingly, in these experiments eosinophil development was seen to ensue when low levels of GATA-1 were expressed whilst higher levels were associated with erythroid and megakaryocytic differentiation. Eosinophil differentiation is prevented by deletion of a double GATA binding site in the GATA-1 promoter (Yu, *et al* 2002a).

### **1.2.6.2 Mast Cells**

Mast cells, although originating from HSCs in the bone marrow, only undergo proliferation and terminal differentiation in connective or mucosal tissues. GATA-1 expression is virtually undetectable in bone marrow mast cell progenitors but is upregulated during maturation in the tissues (Harigae, *et al* 1998). Use of the transgenic “GATA-1 low” mice showed that GATA-1 deficiency caused proliferation of mast cell precursors countered by increased levels of apoptosis, which resulted in normal numbers of mature cells. These mature cells have abnormal morphology and functional defects (Migliaccio, *et al* 2003).

### 1.2.6.3 Dendritic cells

Recently, a new role has been described for GATA-1 in the dendritic cell lineage (Gutierrez, *et al* 2007). Dendritic cells (DCs) can develop from myeloid and lymphoid progenitors. Myeloid derived DCs share precursors with macrophages until late stages of differentiation. GATA-1 is expressed in both myeloid and lymphoid derived DCs and circulating monocytes but not in macrophages. Experiments using a conditional floxed GATA-1 knockout allele (see chapter 4) which can be activated in adult mice, showed that loss of GATA-1 led to a significant reduction in circulating monocytes and dendritic cell numbers but had no effect on macrophages. GATA-1 also appears to be needed for DC activation/maturation in response to inflammatory stimuli (Gutierrez, *et al* 2007). Further investigation of the role of GATA-1 in this lineage is needed.

### 1.2.6.4 Sertoli cells

Finally, GATA-1 expression has been detected in Sertoli cells of the testis in a developmental stage specific pattern. Despite this, no phenotype arises from GATA-1 deletion in this lineage (Lindeboom, *et al* 2003). Its function remains unknown.

## 1.2.7 *GATA-1 also plays roles in lineage commitment*

Although early overexpression studies suggested that different levels of GATA-1 influenced lineage choice in multipotential progenitors (Kulesa, *et al* 1995, Visvader, *et al* 1992), it was unclear whether GATA-1 was acting at the level of lineage commitment or favouring outgrowth of mature precursors. GATA-1 knockout embryos seem capable of production of early pro-erythroblasts and megakaryocytes which then get blocked in their differentiation and display abnormalities in survival pathways (Weiss and Orkin 1995). This led to a general view that GATA-1 was important in maturation and survival but not initial lineage determination. This has recently been challenged (Stachura, *et al* 2006). The association of GATA-1 mutations with the myeloproliferative disorder transient abnormal myelopoiesis (TAM), which displays features of erythroid, megakaryocytic and basophil maturation, suggests activity at a relatively early stage of the haematopoietic hierarchy (Hitzler and Zipursky 2005). In addition GATA1.05 mice are seen to develop both B cell and myeloid leukaemias suggesting that reduced GATA-1 activity affects normal CMP/CLP commitment (Shimizu, *et al* 2004). Further examination of “GATA-1 low” mice reveals a unique progenitor population, not seen in normal mice, capable of trilineage commitment to mast cells, megakaryocytes and erythroid cells (Ghinassi, *et al* 2007). These cells can be induced to return to their normal bipotent meg/erythroid state by ectopic

expression of GATA-1. This suggests GATA-1 also plays a role at the level of MEP/Mast cell progenitor commitment and that levels of this transcription factor determine lineage choice. Finally, studies of thrombopoietin stimulated GATA-1 null ES cells show that these cells can form immature blasts capable of indefinite proliferation. On restoration of GATA-1 activity these cells differentiate into both megakaryocytic and erythroid cells suggesting that they represent a common meg-erythroid progenitor (Stachura, *et al* 2006).

### **1.2.8 GATA-1 isoforms**

As discussed in section 1.2.2, and illustrated in Fig 1.5, GATA-1 appears to have at least two isoforms. The longer isoform GATA-1 full-length (GATA-1FL) represents the normal haematopoietic isoform produced by translation of the full length mRNA product. The short isoform (GATA-1s) lacks the N terminus and has been the subject of intense interest recently due to its association with unique, highly informative, preleukaemic and leukaemic conditions associated with trisomy 21 – transient abnormal myelopoiesis (TAM) and acute megakaryoblastic leukaemia (AMKL). This association is discussed in detail in section 1.3.2. Possible roles of GATA-1s in normal haematopoiesis are discussed below.

#### **1.2.8.1 GATA-1s is produced by alternative translation in mice**

GATA-1s was first detected by western blotting. Antibodies raised against the N terminus of GATA-1 produced a single band with a molecular weight (MW) of 47kD corresponding to GATA1-FL. Use of a polyclonal antibody raised against the zinc-finger domains also produced this 47kD FL band but an additional 40kD band was also seen (Calligaris, *et al* 1995). The GATA-1 gene in mice encodes a unique 1.8kb mRNA with no evidence to date of alternative splicing. Although most eukaryotic mRNAs have a single open reading frame and a single functional translation initiation site, usually the AUG codon that lies closest to their 5' end, it is possible to produce more than one polypeptide from a single mRNA (Kozak 1986a). According to the scanning model of mRNA translation a 40S ribosomal subunit binds to the capped 5' end of mRNA and migrates along the mRNA until it reaches the first AUG codon. At this point it stops, couples with a 60S subunit and initiates protein synthesis (Alberts 2008). Recognition and ribosomal arrest at the first AUG is aided by an optimal surrounding sequence namely CCACCAAUGG (called a Kozak sequence (Kozak 1986b)), the A nucleotide at -3 and the G at +4 being the most critical. If this Kozak sequence is suboptimal some 40S subunits will stop and initiate but others will bypass this site and continue until they reach a second AUG triplet. This “leaky scanning” can therefore produce two protein products from one mRNA – one “long” and one “short”.

GATA-1 translation seems entirely compatible with this model. The first AUG does not possess an optimal Kozak sequence and a second potential translation initiation site occurs in frame at codon 84. Hence leaky scanning would be capable of producing long and short forms of the protein. This was confirmed by in vitro transcription/translation experiments (Calligaris, *et al* 1995). Different methods of isoform production appear to operate in humans - alternative splicing of the mRNA producing an exon 1-2-3 splice for GATA-1FL and exon 1-3 splice variant for GATA-1s has been demonstrated, although it is still possible that alternative translation also operates in this system (Rainis, *et al* 2003).

The original description of GATA-1s was accompanied by expression patterns in the developing embryo (Calligaris, *et al* 1995). The authors looked at lysates from day 8.5 whole embryos and day 11.5 fetal liver. Interestingly, both isoforms were readily detected at day 11.5 but only GATA-1s was seen at day 8.5. This raises the possibility that GATA-1s may play a role in yolk-sac derived haematopoiesis.

#### **1.2.8.2 GATA-1s shows reduced transactivation potential in vitro**

In order to test its functional ability, GATA-1s was co-transfected with a GATA-1 reporter gene into COS cells. These experiments showed an approximately 50% reduction in the ability of GATA-1s to transactivate the reporter gene compared to GATA-1FL, consistent with the notion that the N terminus acts as a transactivation domain binding co-factors to enhance GATA-1 activity (Calligaris, *et al* 1995, Martin and Orkin 1990). Despite this, the identification of interacting partners has remained elusive. Interestingly Runx-1 appears to bind the N and C terminals (Elagib, *et al* 2003). A large phage peptide library screening project identified the HIV accessory protein Nef (Secco, *et al* 2003), although the relevance of this in vivo is unclear. Overall, as discussed below, it has been difficult to demonstrate non-redundant activities for the N-terminus. It is possible that the requirement for this region in non-haematopoietic cell lines may reflect the fact that other essential GATA-1 cofactors are absent in these cells making low affinity interactions with the N terminus rate limiting for transactivation.

#### **1.2.8.3 Creation of a transgenic mouse expressing GATA-1s**

Early experiments looking at the domain function of GATA-1 (discussed in section 1.2.3.3 above) had suggested that an N terminal mutant mimicking the structure of GATA-1s was able to rescue primitive erythropoiesis in GATA1.05 (knockdown) mice. The effect on definitive haematopoiesis was variable depending on expression levels of the mutant – physiological levels were unable to rescue the phenotype whilst overexpression led to



rescue in some mice (Shimizu, *et al* 2001). These experiments contrasted with earlier in vitro work which appeared to show that the N terminal domain was dispensable for erythroid and megakaryocytic differentiation. For example, GATA-1 lacking an N terminal domain was equally effective as N-terminal containing mutants at restoring haematopoiesis during embryoid body formation in GATA-1 deficient ES cells (Blobel, *et al* 1995). In addition, megakaryocytic differentiation can be induced in the absence of the N terminus in the early myeloid cell line 416D (Visvader, *et al* 1995). Finally, rescue of terminal erythroid differentiation in the GATA-1 knockout ES cell line G1E was shown not to depend on the presence of the N terminus (Weiss, *et al* 1997). None of these experiments were able to recapitulate normal patterns of GATA-1 isoform expression and they all only model certain aspects of the role of GATA-1 in developmental haematopoiesis, lineage choice, or terminal differentiation, making their interpretation difficult.

In order to investigate the function of GATA-1s in vivo, a mouse model was created where GATA-1 at the endogenous locus was replaced by a construct capable of exclusive GATA-1s expression (a GATA-1s knock-in model or effectively a GATA-1FL knockout) (Li, *et al* 2005). Many of these mice survived to adulthood and showed essentially normal haematological indices with normal white and red blood cell counts, normal platelet numbers, morphology and function and normal megakaryocyte numbers in adult spleens. However, 10-15% of mice were lost in utero. More detailed examination of the in utero phenotype showed a perturbation of fetal haematopoiesis with hyperproliferation of a transient wave of yolk sac and fetal liver progenitors. At E12.5 affected embryos showed gross fetal liver pallor with reduced numbers of erythrocytes accompanied by a large increase in the number of cells expressing the megakaryocytic marker CD41. These CD41<sup>+</sup> progenitors displayed marked hyperproliferation and reduced, but not blocked, maturation when cultured in vitro. These abnormalities were first detected at E9.5 in the yolk sac and the abnormal cell population appeared to have shifted to fetal liver by E11.5, however expression here was only transient with a falling off of progenitor numbers by E14.5 and essentially normal appearances by E16.5. The effect of GATA-1s on gene expression during megakaryocytic differentiation was then assessed by microarray profiling of GATA-1s and wild type GATA-1 fetal liver-derived megakaryocytes. An additional comparison was also made with the same cells derived from mice carrying the megakaryocyte specific GATA-1 knockdown mutation  $\Delta$ Neo $\Delta$ HS discussed in section 1.2.5.1 above. Overall this analysis shows that megakaryocytic genes, normally upregulated on differentiation, are still upregulated in the presence of GATA-1s but to a lesser extent. This contrasts with  $\Delta$ Neo $\Delta$ HS where mature gene expression is blocked. In contrast, GATA-1s fails to downregulate a number of key genes normally repressed by

GATA-1. These include myc, myb, GATA-2, PU.1 and Ikaros. The first three genes are important for proliferation of progenitors whilst the last two play essential roles in lineage choice (myeloid and lymphoid respectively). Overall this data suggests that GATA-1s is capable of uncoupling differentiation and proliferation and that, far from being an essential transactivation domain, the N terminus may actually function as a critical repressor of proliferative genes and those responsible for alternative lineage choice. One important limitation of this study is that, although knocked-in to the endogenous locus, GATA-1s expression does not mimic its physiologic expression pattern. GATA-1s will be produced in place of GATA-1FL as the first ATG is followed by the GATA-1s coding sequence. Interestingly a second mouse model has been created using N-ethyl-N-nitrosurea mutagenesis that carries a mutation in the initiation codon of GATA-1 thereby preventing translation of GATA-FL. In contrast to the GATA-1s knock-in this mutation is embryonic lethal in males at a time equivalent to GATA-1 null mice. Heterozygous females exhibit thrombocytopenia and a fetal blast population with mast cell like progenitor characteristics. The authors were unable to document GATA-1s production at the protein level despite the presence of mRNA transcripts (Majewski, *et al* 2006).

#### **1.2.8.4 Insights from a naturally occurring human mutation**

The discovery of a Brazilian family with an inherited mutation in GATA-1FL which preserves GATA-1s function has actually produced a far better knockout model than those described above! In this family a mutation in the splice site at the end of exon 2 means that no FL GATA-1 transcript is produced but that GATA-1s is still transcribed at a level equivalent to unaffected family members (Hollanda, *et al* 2006). Analysis of the family uncovered 7 affected males over 2 generations with full penetrance. Female heterozygotes had no detectable phenotype despite the prediction that they would express the mutant allele in 50% of their cells due to Lyonisation. All affected males had a moderate to severe macrocytic anaemia, accompanied by neutropenia but normal platelet counts. Bone marrow examination revealed trilineage dysplasia, reduced erythropoiesis and granulopoiesis but normal or increased numbers of megakaryocytes some of which were small and immature. Despite normal counts, peripheral blood platelets were morphologically and functionally abnormal. Overall these data suggest that GATA-1s is incapable of supporting normal adult erythropoiesis and megakaryopoiesis. Additionally, neutropenia has not previously been described in association with GATA-1 deficiency, the mechanism for this is unclear. Interestingly this kindred- and the mouse GATA-1s knock-in described above- do not appear to have an increased incidence of leukaemia or transient abnormal myelopoiesis - discussed below. The differences between the human and mouse

data may be explained by differences in levels of expression of GATA-1s (high in the mouse, physiological in the human). It was previously shown that high levels of a GATA-1 NT mutant were required to rescue definitive haematopoiesis (Shimizu, *et al* 2001).

#### **1.2.8.5 Summary of known GATA-1s function**

Overall these studies, although sometimes conflicting, suggest that GATA-1s may have a role in normal yolk-sac haematopoiesis but that the expression of GATA-1FL is required for normal levels of definitive haematopoiesis (Hollanda, *et al* 2006, Shimizu, *et al* 2001). Mutational analysis has shown a vital role for the N-terminus in regulation of fetal liver megakaryocyte proliferation (Kuhl, *et al* 2005, Muntean and Crispino 2005), and microarray studies have confirmed inadequate repression of genes involved in cellular proliferation and lineage choice in GATA-1FL knockouts (Li, *et al* 2005). A model could be proposed whereby GATA1s is responsible for a background level of megakaryopoiesis and erythropoiesis which may be particularly important in early embryogenesis when early reliable establishment of platelet and red blood cell production is an essential requirement for embryonic survival. The N-terminus (GATA-1FL) is then required to take part in sophisticated transcriptional networks associated with lineage choice and maintenance of a proper balance between gene repression and activation in definitive haematopoiesis. Whether GATA-1s has any unique role in haematopoiesis, or whether it is redundant in the presence of a fully functioning GATA-1FL allele, has not been tested to date. The role of GATA-1 mutations in the development of leukaemia will be discussed below.

### **1.3 Leukaemia**

Leukaemia is the medical term given to malignancies of the haematopoietic system. Study of leukaemia has shed much light on the genetic basis for haematopoiesis and knowledge of key molecular events in haematopoiesis has allowed some major advances in the management of patients with leukaemia. This disease is a good example of the interdependence between basic biology and translational medicine. Leukaemia is classified according to the presumptive cell of origin and kinetics of disease (acute or chronic). The major broad classification of acute leukaemias is into acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL). AML is further subdivided using a number of different systems. Traditionally sub-classification of AML was based on cell of origin using an agreed nomenclature known as the FAB (French-American-British) class i.e. undifferentiated myeloblast, promyelocyte, myelomonocyte, monocyte, eosinophil, basophil, erythroid, or megakaryocyte. More recently the classification has been adapted

by the World Health Organisation to reflect differences in disease biology and to incorporate key cytogenetic features that determine outcome (Jaffe 2001). ALL is subdivided into B cell ALL and T cell ALL, there is then further subdivision according to the apparent stage of developmental arrest e.g. pro-B, preB, mature B.

The causes of leukaemia are generally unknown. There are few well-established risk factors with the exceptions being; exposure to ionising radiation, chemicals such as benzene or previous chemotherapy (Hoffbrand, *et al* 2005). The majority of cases are sporadic, although occasionally acute myeloid leukaemia can arise secondary to an inherited bone marrow failure syndrome (Alter 2002) or a germline mutation in a haematopoietic regulatory gene e.g. Runx-1 (Owen, *et al* 2008).

### **1.3.1 The genetic basis for leukaemia development**

Leukaemias are characterised by deregulation of proliferation, maturation and cell survival/apoptosis pathways with the balance shifted towards survival even in the face of cellular stress. Initial clues as to the molecular basis for leukaemia came from the study of recurring chromosomal translocations identified in subsets of leukaemic cells. Cloning and characterisation of these translocation breakpoints led to identification of key genes whose disruption could lead to a malignant phenotype. The number of identified translocations has rapidly increased with over 100 now described (Gilliland 2001). This far exceeds the number of identified subtypes of leukaemia. In addition it has become clear that many other genetic changes occur in leukaemia cells such as point mutations in genes or their regulatory regions, intrachromosomal amplification, loss of heterozygosity, epigenetic histone modifications, alterations in protein degradation and changes in microRNA profiles. Some of these abnormalities are primary transforming events whilst others are secondary to genomic instability and disrupted regulatory processes in transformed cells, or may follow selective pressure such as chemotherapy treatment. Despite this apparent complexity a number of themes emerge that help us to understand the molecular basis for leukaemia development. These are discussed below:

#### **1.3.1.1 Class 1 and Class 2 mutations**

Very early observations on the nature of malignant transformation and the kinetics of disease acquisition, well before the era of molecular medicine, suggested that the development of cancer requires more than one genetic change (Ashley 1969). This eventually led to the now well established concepts of cellular proto-oncogenes and tumour

suppressor genes (Knudson 2001). Study of the myriad of translocation partners and other genetic mutations in leukaemias has identified that genes often fall in to one of two groups which are now termed Class 1 and Class 2. Class 1 mutations confer a proliferative or survival advantage to cells but do not affect haematopoietic differentiation, these often involve signal transduction pathways or genes with roles in the cell cycle, examples include the bcr/abl translocation in CML or Flt-3 activating mutations in AML. Class 2 mutations primarily affect differentiation and have only modest effects on proliferation/survival, these often involve transcription factors, for example the PML:RAR $\alpha$  translocation in acute promyelocytic leukaemia (Gilliland 2001). It is noteworthy that transcription factor mutations are rarely null mutations instead they usually produce either quantitative or qualitative effects on function, basal levels of transcription factor activity are probably crucial for cell survival (Rosenbauer, *et al* 2005). It is hypothesised that, in order to undergo leukaemic transformation, a cell must acquire both a Class 1 and a Class 2 mutation (Gilliland 2001). This hypothesis was originally questioned by the observation that many acute myeloid leukaemias only appeared to carry a single dominant mutation in a class 2 gene such as the t(8:21) – AML-1-ETO or t(15:17) PML-RAR $\alpha$ , but more recently it has emerged that these cells often carry additional mutations in class I genes that are not detectable by conventional cytogenetics such as a Flt-3 activating mutation (Nakao, *et al* 1996). In fact activation of multiple signal transduction pathways appears to be a feature of AML and clinical outcome is inversely related to the number of pathways activated (Kornblau, *et al* 2006). One conclusion from these observations is that a single “targeted therapy” is unlikely to be sufficient for the vast majority of leukaemias. Instead an approach targeting both Class 1 and Class 2 mutations is much more likely to be successful. Another prediction is that in some cases it should be possible to identify premalignant clones that have acquired one class of mutation but are yet to acquire a second. This is discussed below:

### **1.3.1.2 In utero acquisition of mutations**

One observation that led to the two hit hypothesis of malignancy was that the incidence of malignancy generally rises with age, presumably reflecting the time needed to acquire the requisite number of mutations. An exception to this general rule is seen in childhood malignancy. Tumours in childhood often show very distinct age profiles. This may reflect transformation of a developmentally restricted progenitor population. These fetal or neonatal progenitors often have periods of rapid proliferation or other cellular activity that puts them at increased risk of mutations. Any cell acquiring such a mutation has a limited window in which to acquire a second mutation before the population as a whole terminally

differentiates or is extinguished. Examples of such tumours include retinoblastoma – involving fetal retinoblasts and neuroblastoma involving embryonic neural crest cells.

In the haematopoietic system certain leukaemias show distinct peaks of incidence in childhood including most types of ALL and certain rare myeloid leukaemias. The first indication that a prenatal event may be involved came from observations of leukaemia in identical twins. The concordance rate for leukaemia in monozygotic twins varies greatly depending on the subtype e.g. approximately 50% for infant leukaemias, 10% for common childhood ALL and <1% for adult ALL or AML (Greaves 2005). This was initially interpreted to reflect a common inherited susceptibility to infant leukaemia. However further investigation has shown that this high rate of concordance is due to blood cell chimaerism between twins that share a common placental circulation (mono-chorionic twins). Using molecular techniques it has been possible to show that the twins share identical breakpoints in leukaemic fusion genes. This implies that the fusion arose in a haematopoietic progenitor in one twin and progeny of this cell passed via the placental circulation to the other twin. The need for a second postnatal event means that concordance is not 100%. In fact a concordance of 50% makes a second postnatal (or independent pre-natal) event extremely common. This may be because the majority of infant ALLs are associated with MLL gene abnormalities and it has recently been shown that this mutation renders the cell very vulnerable to DNA damage (Eguchi, *et al* 2006). For childhood common ALL concordance is still high. Identification of twins discordant for common ALL has allowed elegant studies looking at the kinetics of pre-malignant clones in an unaffected twin who carries a Class 2 mutation but is yet to develop acute leukaemia (and may never do so) (Hong, *et al* 2008). These findings have been shown to be of general relevance to childhood leukaemia. In particular the startling observation has been made that 1% of a randomly selected sample of newborn blood spots (taken as part of routine neonatal Guthrie screening for metabolic diseases) have a detectable Tel:AML-1 t(12:21) translocation (Mori, *et al* 2002). Fortunately, only about 1% of these children will go on to develop ALL providing further support for the need for more than one mutation. These observations have also helped develop hypotheses for the likely causes of the peak of leukaemia incidence in children aged 2-5 years old. In particular it hypothesised that prenatally the rapid proliferation of lymphoid precursors accompanied by physiological gene rearrangements needed to produce unique B and T cell antigen recognition receptors predisposes to chromosomal translocations. If these translocations lead to a proliferative or survival advantage then these cells will undergo clonal expansion of a usually rare intermediate population. Postnatally this small clone remains static in size and would normally be extinguished unless a second selective pressure is applied. The nature of this

selective pressure is unknown but evidence is accumulating that a delayed exposure to certain infections may be responsible (Greaves 2006).

Although the majority of evidence for the in utero origins of leukaemia has arisen from study of childhood ALL there is also evidence for a similar process in some forms of childhood AML. In one study Guthrie cards were retrieved from children presenting with AML and the t(8:21) translocation – AML1-ETO, 50% of these children had the same mutation detectable at birth with the oldest child being 12 years 2 months of age when they presented with overt AML (Wiemels, *et al* 2002). This data is supported by mouse models of AML-1-ETO which do not develop leukaemia unless exposed to low dose chemical carcinogens – presumably triggering a second mutation (Higuchi, *et al* 2002).

### **1.3.1.3 The cellular origins of leukaemia**

As alluded to above, leukaemic cells often have normal cellular counterparts (Greaves 1986). An emerging concept in cancer biology is that tumours may contain some cancer stem cells - rare cells with indefinite self-renewal properties, that drive the formation of tumours (Reya, *et al* 2001). The existence of leukaemia stem cells was first suggested by experiments in the 1960s and 70s that showed that only a small subset of leukaemic cells were capable of extensive proliferation in vitro. Also, in vivo transfer experiments were only able to demonstrate spleen colonies in 1-4% of mice (Bruce and Van Der Gaag 1963). When considering the cell of origin for the leukaemic transformation in AML two possibilities exist. Firstly, leukaemic transformation may occur in the normal HSC or secondly a more committed progenitor (or even a mature cell) could re-acquire self renewal properties secondary to transformation. HSCs are particularly good targets for leukaemic transformation as maintenance of self-renewal activity may be easier than turning it on de novo. In addition HSCs are long lived cells which undergo multiple rounds of cell division giving them ample time and opportunity to acquire multiple mutations. This does not mean that restricted progenitors cannot be the target cells for leukaemic transformation – they could inherit an original first hit from an HSC and the second hit could be acquired in a downstream progenitor. Alternatively, a committed progenitor could acquire a mutation that conferred stem cell like properties of self renewal. Evidence for both pathways exists, although it is likely that HSC transformation is commoner. One elegant study showed that in most AML subtypes transfer of a primitive CD34+CD38- subpopulation (comprising only 0.2-1.0% of the patients leukaemic cells) was capable of transferring AML into NOD/SCID mice, whilst the CD34+CD38+ cells which comprised the bulk of the tumour did not sustain engraftment (Bonnet and Dick 1997). These

CD34<sup>+</sup>CD38<sup>-</sup> cells did not only sustain stable leukaemic engraftment but also produced a leukaemia that was morphologically and immunophenotypically identical to the original patient's leukaemia. This implies that this primitive population of leukaemic stem cells were capable of partially differentiating to produce more mature although still aberrant progeny – a leukaemic cellular hierarchy exists that parallels the normal haematopoietic cellular hierarchy. This leads to the important conclusion that the bulk cell population in the tumour does not necessarily equate to the cell of origin of the leukaemia. Evidence does exist for the alternative model of acquisition of stem cell properties in more mature cells. In particular transfection of a leukaemic fusion gene t(11:19) MLL-ENL transcript into either HSCs, common myeloid progenitors or granulocyte-macrophage progenitors produces exactly the same leukaemia, irrespective of progenitor stage, on transfer to mice (Cozzio, *et al* 2003). In the case of ALL the concept of a stem cell population is less well established. This probably reflects the unique properties of lymphocytes which allow clonal expansion at mature stages of differentiation, hence the ability to repopulate NOD/SCID mice has been demonstrated in many different fractions of the leukaemic cell population, including relatively mature CD19<sup>+</sup> compartments (le Viseur, *et al* 2008). Overall, the concept of leukaemic stem cells, particularly in AML, is an important one. Therapies directed at the bulk cells in the tumour will produce rapid cytoreduction (and therefore an apparent response) but may leave the cancer stem cell population untouched. This will inevitably lead to relapse at a later date. Understanding the properties of the cancer stem cell will be crucial for effective targeted therapy. Unfortunately, akin to normal HSCs, despite being capable of expansion and self renewal these cells are often relatively quiescent, and may be relatively drug resistant due to expression of high levels of efflux pumps (Elrick, *et al* 2005). This makes devising effective therapy particularly challenging.

#### **1.3.1.4 Primary and secondary changes in leukaemic cells**

Despite the extensive acquisition of information on molecular events in leukaemic cells we are still far away from a comprehensive explanation of the stepwise acquisition of abnormalities that lead to a fully leukaemic phenotype. One of the major problems is that the vast majority of leukaemias only present when fully established and causing clinical symptoms. At this stage the cells have often acquired multiple genetic abnormalities some fundamental to leukaemogenesis and some secondary events. Dissecting out the primary initiating events can be very difficult at this stage. By analogy with the concept of leukaemic stem cells discussed above, directing therapies at secondary events will leave the primary initiating clone untouched. It is vital to understand the early steps if therapies



are to be effective. In addition, using this information it may be possible to identify leukaemia at a premalignant stage. This raises the possibility of preventative treatment rather than waiting for a fully malignant, clinically aggressive phenotype, possibly associated with a reduced chance of cure. This is akin to screening for, and treating, premalignant lesions in other systems such as colonic polyps or cervical intraepithelial neoplasia. The existence of clinically rare leukaemias with a recognised pre-leukaemic phase has provided a unique opportunity to study the multiple genetic and cellular events needed to produce full-blown leukaemia *in vivo*. The paradigm for this is Down syndrome associated acute megakaryoblastic leukaemia discussed below.

### **1.3.2 Down syndrome and leukaemia**

Down syndrome is caused by Trisomy 21 and has an incidence of 1 in 600-800 live births making it the commonest human chromosomal abnormality. It has long been known that children with Down Syndrome have an increased incidence of leukaemia approximately 10-20 fold above that of the general population (Lange, *et al* 1998). This is not due to a general cancer predisposition syndrome, the incidence of other tumours is often reduced below background (Hasle, *et al* 2000). Both ALL and AML are commoner in these children but it is the existence of two rare, interrelated, and almost unique, phenomena that has been particularly informative- namely a transient abnormal myelopoiesis (TAM) and acute megakaryoblastic leukaemia (AMKL).

#### **1.3.2.1 TAM and AMKL - natural history and clinical features**

Children with Down syndrome (DS) have a 500 fold increased incidence of acute megakaryoblastic leukaemia which nearly always manifests between the ages of 1 and 4 years (Lange, *et al* 1998). In fact children presenting in this age range with AMKL almost always have evidence of trisomy 21 – if not already clinically apparent then genetic testing often reveals mosaicism or, more rarely, an acquired trisomy 21 in the leukaemic blast cell population. This AMKL is clinically and biologically distinct from an alternative infant AMKL which is associated with a (1:22) translocation. The infant form has no association with Down syndrome and is clinically aggressive with a poor treatment outcome (Ma, *et al* 2001). In contrast DS associated AMKL has a good prognosis in children, with 10 year overall survival rates of 79% (Hama, *et al* 2008). This may reflect the presence of a number of genes on chromosome 21 that enhance sensitivity to chemotherapeutic agents

such as the cystathionine- $\beta$ -synthetase which potentiates the effects of cytarabine – a drug known to be very effective in AML (Taub, *et al* 1999).

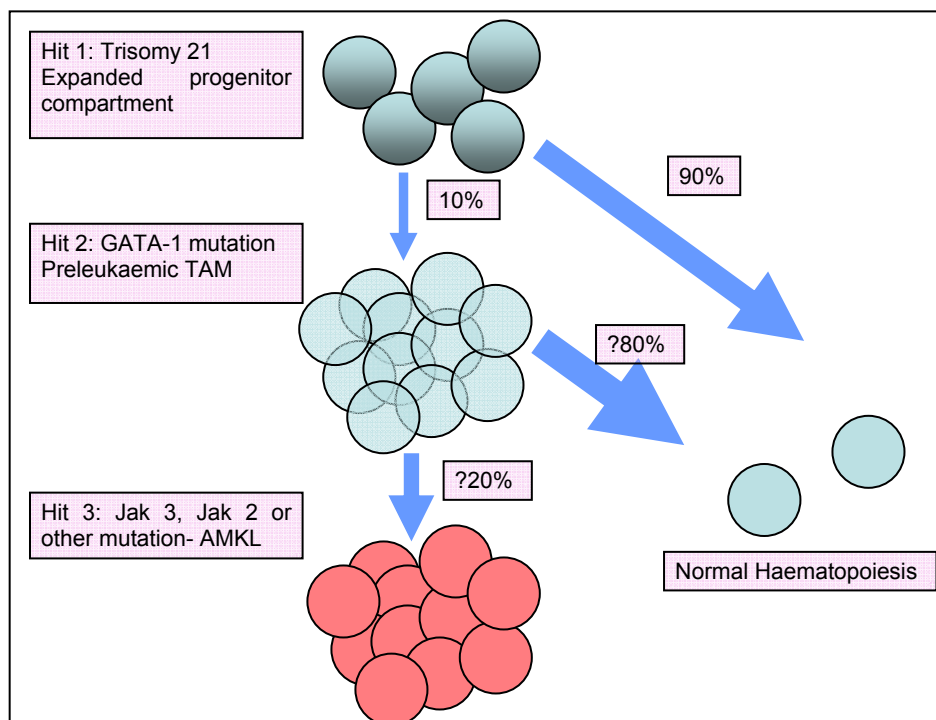
Children with Down syndrome display another relatively unique haematological disorder known variously as transient abnormal myelopoiesis (TAM), transient myeloproliferative disorder (TMD) or transient leukaemia. As the name suggests this disorder is associated with increased numbers of immature myeloid cells in the circulation which are often present at or before birth and usually resolve within the first three months of life. The classification of this disorder has caused some controversy as these cells are phenotypically indistinguishable from leukaemia cells – they are morphologically and immunophenotypically identical to AMKL cells, monoclonal (Kurahashi, *et al* 1991) and may have acquired additional genetic mutations (Zipursky 2003). This would suggest that they are leukaemic blasts. Evidence against this comes from their biological behaviour – these cells often spontaneously disappear without any active treatment. Further elucidation of the molecular basis of TAM and full-blown AMKL, discussed below, has led to the conclusion that TAM represents a preleukaemic phase of AMKL. Like AMKL, TAM is also tightly linked to the presence of trisomy 21. Virtually no cases have been described in otherwise normal children.

Clinically TAM is often asymptomatic and found incidentally on a routine full blood count performed for other indications such as cardiac surgery. However, it can present clinically with organomegaly or a skin rash. In some patients (reported to be up to 20% in one prospective study (Al-Kasim, *et al* 2002)) the disease may take a more severe course with hepatic failure or cardiopulmonary disease. The main pathology seems to be extensive blast cell proliferation in the liver with associated hepatic fibrosis. This can cause significant morbidity and mortality but may respond to low dose chemotherapy (cytarabine) if treatment is started early enough (Al-Kasim, *et al* 2002). The incidence of TAM is currently unknown. This is because children with Down syndrome do not all undergo full blood count analysis at birth. Since TAM is often clinically silent this means a large number of cases go undetected. A number of small cohort studies have put the incidence at about 10% of all babies with Down syndrome (Zipursky, *et al* 1997). A recent prospective study reported the outcome for 48 children with TAM, this confirmed the early mortality in severe cases but also that 89% of children spontaneously cleared their blast population. 19% of infants went on to develop AMKL. Risk factors for progression to AMKL were male sex and presence of additional cytogenetic abnormalities at diagnosis of TAM (Massey, *et al* 2006). It is worth noting that the various figures in the literature are inconsistent. For example, the risk of AMKL in DS is quoted as approximately 1 in 500,

the risk of TAM 10% and the risk of progression from TAM to AMKL 20-30%. If this latter figure were correct then the risk of AMKL would be 2-3 in 100, not 1 in 500. This probably reflects the fact that there is a reporting bias – the incidence of AMKL in patients presenting to tertiary referral centres with Down syndrome complications may well be higher than in an unselected population. A number of cohort studies are currently assessing the true incidence and natural history of TAM (Vyas and Roberts 2006).

From these basic clinical facts, a number of predictions can be made. Firstly the development of AMKL must require at least 3 genetic events; trisomy 21 is required for TAM but not sufficient, as 90% of Down syndrome children appear haematologically normal; a second event is needed to develop TAM but again this is not sufficient for AMKL as most children recover completely; and a third event is then needed in children who go on to develop AMKL, this stepwise progression is illustrated in Fig 1.7. A second prediction is that the age range of children with AMKL suggests leukaemic transformation occurs in a cell with developmentally restricted potential such as a yolk sac or fetal liver progenitor. Lastly, the nature of the blasts in TAM and AMKL suggest involvement of genes important in the specification of megakaryocytic and erythroid lineages.

Over the last 6 years remarkable progress has been made in unravelling the molecular basis of TAM and AMKL. This will be reviewed below:



**Fig 1.7 At least 3 hits are needed to develop AMKL**

### 1.3.2.2 GATA-1 mutations and DS TAM/AMKL

The tight association between trisomy 21 and TAM/AMKL led to the assumption that any unique genetic mutations in this disorder would be discovered in chromosome 21 genes. It was surprising therefore, that the real breakthrough came with the discovery in 2002 that all (6/6) cases of DS-AMKL studied had mutations in the X-linked gene GATA-1. This association between GATA-1 mutations and Down syndrome AMKL was tightly linked – other DS ALL and AML, other non-DS AMKL and various AML, MDS and healthy control samples (92 samples in total) were all negative for GATA-1 mutations (Wechsler, *et al* 2002). The detected GATA-1 mutations all resulted in premature termination of GATA-1 in the N terminal activation domain i.e. they prevented synthesis of GATA-1FL but should not interfere with production of the GATA-1s isoform. These predictions were confirmed by western blotting showing absence of the 47kDa GATA-1FL protein but presence of a 40kDa GATA-1s band. The mutations were not present in remission bone marrow samples from the same patients, confirming that they were somatically acquired. This report was confirmed by a number of groups all of whom reported the extremely high incidence of GATA-1 mutations in DS AMKL and the clustering of these mutations in the N-terminus. Overall 45 out of 53 reported cases have confirmed GATA-1 mutations (Ahmed, *et al* 2004, Groet, *et al* 2003, Hitzler, *et al* 2003, Magalhaes, *et al* 2006, Wechsler, *et al* 2002, Xu, *et al* 2003). To date only one case has been reported of a GATA-1 mutation in an individual without trisomy 21- an adult with AMKL (Harigae, *et al* 2004). This represents a virtually unprecedented level of concordance between a genetic mutation, an underlying constitutional disorder and risk of leukaemia.

An obvious question is at what stage GATA-1 mutations exert their effect – are they responsible for the eventual leukaemic transformation or do they produce the preleukaemic disorder TAM? A number of studies addressed this and confirmed that the latter was correct – GATA-1 mutations in the N-terminus are also seen in virtually all cases of TAM. 73 out of 80 reported cases have detectable GATA-1 mutations (Ahmed, *et al* 2004, Groet, *et al* 2003, Hitzler, *et al* 2003, Magalhaes, *et al* 2006, Rainis, *et al* 2003, Shimada, *et al* 2004, Xu, *et al* 2003). These studies also established a number of other important facts:

***The GATA-1 mutation is acquired prenatally:*** A number of observations point to the prenatal acquisition of GATA-1 mutations. TAM is usually present at birth and Guthrie blood samples taken in the first 2 weeks of life have shown GATA-1 mutations (Ahmed, *et al* 2004). In addition two pairs of identical twins have been reported with TAM. In both cases the two babies carried identical GATA-1 mutations implying in utero acquisition of

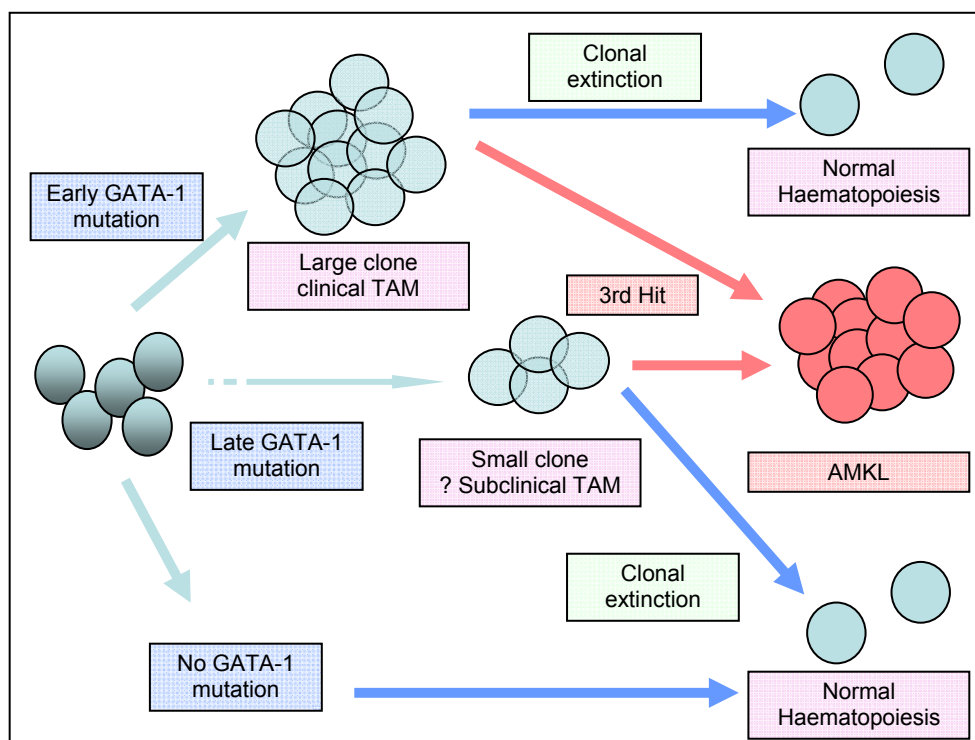
the mutation in one twin which was then passed on to the other twin via a shared circulation (Rainis, *et al* 2003, Shimada, *et al* 2004). There are many case reports of Down Syndrome fetuses presenting before birth with hydrops fetalis or massive hepatomegaly detected on ultrasound scanning – clinical features of TAM have been recorded on fetal blood sampling and, in one study, a GATA-1 mutation was confirmed in a stillborn infant with clinical signs of TAM (Heald, *et al* 2007). Fetuses with TAM like features have been reported from 25 weeks gestation onwards (Robertson, *et al* 2003). Finally, a study using autopsy material from fetuses with Down syndrome found GATA-1 mutations in 2/9 fetal liver samples at 21 and 23 weeks gestation. These fetuses were not known to have features of TAM (Taub, *et al* 2004). Although a very small study, this level of detection suggests that the incidence of GATA-1 mutations may be larger than that of clinically detectable TAM.

***GATA-1 mutations in TAM and AMKL in the same patient are identical:*** In cases where samples from TAM and AMKL have been available in the same patient the mutation in GATA-1 has always been identical (Rainis, *et al* 2003). In addition in patients presenting with AMKL with no obvious antecedent history of TAM, examination of Guthrie spots has confirmed a GATA-1 mutation identical to the AMKL associated mutation in 3 out of 4 cases studied (Ahmed, *et al* 2004). This provides molecular evidence for the long held assertion that AMKL arises due to evolution of a persisting TAM clone rather than as a completely independent event i.e. that TAM represents a preleukaemic phase of AMKL. This same study showed that the size of the GATA-1 mutation clone can vary considerably at birth. This leads to a hypothesis that the clinical presentation of TAM is determined by clone size, which may in turn be related to the gestational age at which the GATA-1 mutation was first acquired (Vyas and Roberts 2006). This is illustrated in fig 1.8.

***In some patients multiple GATA-1 mutations are detectable:*** Using sensitive techniques it is possible to detect, in some cases of TAM, more than one mutant clone, each carrying a different GATA-1 mutation, (Ahmed, *et al* 2004). This implies that either trisomy 21 is associated with very high rates of somatic mutation (unlikely due to the specificity of GATA-1 mutations and the lack of mutations in other genes studied such as Runx-1 (Xu, *et al* 2003)) or, as is thought more likely, the GATA-1s mutation confers a selective advantage to progenitor cells. The clones may vary in size although the size of the clone does not seem to determine the risk of progression to AMKL (Xu, *et al* 2006).

***GATA-1 mutations and trisomy 21 are only required in the haematopoietic cells for establishment of TAM/AMKL:*** Occasional cases exist of a TAM phenotype in patients

who do not appear clinically to have Down syndrome. These patients usually have either mosaicism for trisomy 21 or have an acquired trisomy 21 in their blast cell population. One report of identical twins with just such an acquired trisomy 21 has shown clinical presence of TAM in both twins and progression to AMKL in one twin. Trisomy 21 was not detectable in karyotypic analysis of non-TMD cells and disappeared with disease resolution (Carpenter, *et al* 2005). Since these patients displayed the full TAM/AMKL phenotype it appears that both trisomy 21 and GATA-1 mutations are only required in the haematopoietic cells. This suggests that these genetic abnormalities act in a cell autonomous manner rather than altering the microenvironment or extrinsic signalling pathways to cause excessive cellular proliferation.



**Fig 1.8 Proposed impact of clone size and acquisition of mutations on clinical presentation**

During Down syndrome fetal haematopoiesis three different scenarios may occur. 1) Early acquisition of a GATA-1 mutation may lead to a large abnormal clone which may present in the pre- or post-natal period as clinically detectable transient abnormal myelopoiesis (TAM) 2) A late GATA-1 mutation may produce a small clone resulting in sub-clinical TAM 3) no GATA-1 mutation occurs, producing normal haematopoiesis. Both large and small clones may either undergo clonal extinction, resulting in normal haematopoiesis, or acquire additional genetic changes (a 3rd hit) to produce a full-blown acute megakaryoblastic leukaemia (AMKL).

***GATA-1 mutations are seen in Guthrie spots taken from DS children with no clinical evidence of TAM:*** A retrospective study of 21 healthy Down syndrome neonates with no clinical evidence of TAM showed GATA-1 mutations in 2/21 of their Guthrie spot neonatal blood samples (Ahmed, *et al* 2004). Another larger study looked at 585 Guthrie

tests from children with trisomy 21 and identified GATA-1 mutations in 22 of them (3.8%) (Pine, *et al* 2007). This leads to questions about the clinical definition of TAM – should all children with GATA-1 mutations be considered to have TAM or only those with clinically evident disease? Interestingly, it appears that clinical evidence of TAM (presumably reflecting a large abnormal clone size) is not associated with an increased risk of transformation to AMKL and in fact these patients often respond better to treatment than patients with no antecedent history of TAM (Klusmann, *et al* 2008). These are important considerations as parents of children with TAM are obviously concerned about the risk of progression to AMKL – this risk is likely to vary depending on the definition of TAM used. Prospective studies looking at the natural history of TAM are required.

***Although alternative translational start points are identifiable, only GATA-1s protein is ever detectable in these cells:*** Over 100 GATA-1 mutations have been described to date all of which fall in the N terminal domain. Some of these mutations lead to predictions of alternative, out of frame, start codon usage between the codon 1 and codon 84 – the start codon for GATA-1s (Hitzler, *et al* 2003). Many studies have not shown actual evidence of GATA-1s production by western blotting, but in those that have done so, only the 40kDa GATA-1s band has ever been detected, even if other start codons appear possible (Xu, *et al* 2003). This implies that the GATA-1s protein is vital for the pathology of the disease rather than just loss of the full length isoform.

### **1.3.2.3 Evidence for a fetal liver progenitor cell population as the target for GATA-1 mutations**

A number of pieces of evidence point to TAM/AMKL arising from transformation of a fetal haematopoietic progenitor rather than definitive adult haematopoietic cells. Firstly TAM and AMKL are characterised by blast accumulation in the liver and/or spleen with relative sparing of the bone marrow (Miyauchi, *et al* 1992, Zipursky 2003). Blasts from TAM show evidence of fetal and adult globin expression (Ito, *et al* 1995). As discussed in section 1.2.8.3 the GATA-1s knock-in mouse shows a specific developmentally restricted defect in yolk-sac and fetal liver haematopoiesis which is then followed by normal adult haematopoiesis (Li, *et al* 2005). Most compellingly however is the simple observation that TAM and AMKL are developmentally restricted to the first 3 months and 4 years of life respectively. AML in children with Down syndrome older than 4 usually lacks the GATA-1 mutation and biologically is more akin to sporadic AML (Hasle, *et al* 2008). Adults with acquired trisomy 21 and AMKL do not have detectable GATA-1 mutations and their disease is biologically distinct from DS-AMKL (Vyas and Roberts 2006). The regression

of TAM corresponds to the switch from fetal liver to bone marrow haematopoiesis (Crispino 2005). This age restriction would be compatible with the transformed cell having an intrinsically limited lifespan or with a failure of the postnatal haematopoietic environment to sustain embryonic precursors. Although it should be noted that an alternative explanation is that the GATA-1s containing clones have a proliferative advantage, but are not immortalised, and therefore eventually die out due to failure of self-renewal. Of interest, in one study, telomerase activity (associated with malignancy and the ability of cells to undergo multiple rounds of division without senescence) was found in 15/29 samples from AMKL patients but only 4/34 cases of TAM. In addition 3 out of the 4 telomerase positive TAM cases were associated with particularly severe disease (Holt, *et al* 2002). In terms of the lineage affiliation of the progenitor cell in AMKL, the blasts in TAM and AMKL express both erythroid and megakaryocytic genes at the mRNA level and display typical erythroid and megakaryocytic cell surface marker expression patterns (Ito, *et al* 1995, Miyauchi, *et al* 1992). This suggests the involvement of a cell capable of at least a bipotential meg-erythroid fate. Interestingly many cases of TAM are also associated with increased numbers of circulating basophils and more rarely eosinophils. In vitro culture of TAM blasts also sometimes produces basophils (Zipursky 2003). The origin and commitment of cells to the basophil/mast cell lineage remains one of the unresolved features of haematopoiesis, as does the developmental relationship between the two lineages (Arinobu, *et al* 2005, Gurish and Boyce 2006). The recent description of a tripotential meg-erythroid-mast cell progenitor population in “GATA-1 low” mice may be relevant here (Ghinassi, *et al* 2007). The discovery of yolk-sac derived primitive megakaryopoiesis discussed in section 1.1.1.2 (Tober, *et al* 2007) provides another possible target cell for GATA-1s mutations.

Overall, the evidence above suggests that the GATA-1 mutation is acquired in fetal life and is responsible, in conjunction with trisomy 21, for development of the preleukaemic condition TAM. GATA-1s therefore acts as a second genetic hit in this condition (see Fig 1.7). What have we learnt about the first and third hits?

#### **1.3.2.4 First genetic hit – Trisomy 21**

Perhaps least is known about the contribution of trisomy 21 to the development of AMKL. Trisomy 21 appears to be an essential pre-requisite for the oncogenic activity of GATA-1 mutations, as shown by the absence of leukaemia in mouse models of GATA-1s and the human GATA-1s expressing kindred (Hollanda, *et al* 2006, Li, *et al* 2005). Trisomy 21 itself seems to predispose to haematopoietic disorders independent of the increase in



leukaemia. Children with Down syndrome often display polycythaemia, macrocytosis and thrombocytosis at birth (Kivivuori, *et al* 1996). In addition, study of trisomy 21 is of general relevance to leukaemia in the non-DS population –acquired tri or tetrasomy 21 is a frequent finding in ALL and internal amplification of chromosome 21 is a very poor prognostic feature in these disorders. It seems possible that trisomy 21 may exert its effect due to a gene dosage with overexpression of an oncogenic gene located on chromosome 21 co-operating with GATA-1 mutations to produce TAM/AMKL. Consistent with this, mouse models of Down syndrome show elevated expression of most triplicated genes, to about 1.5 times normal, across a wide range of tissues throughout development (Roper and Reeves 2006). However, it is now becoming clear that trisomy 21 may exert its effects in many ways other than a pure gene dosage effect. Paradoxically, the more that is discovered about the genes that are responsible for the Down syndrome phenotype (including recent description of a key role for the NFAT transcription factors that recapitulate many features of Down syndrome when over-expressed in mice (Arron, *et al* 2006, Gwack, *et al* 2006)) the more complex the picture becomes. Besides simple gene dosage, the effects of trisomy 21 may be caused by altered balance between recessive and dominant alleles or many coincident small effects. In addition interaction between trisomic gene products and normal disomic genes in the cell will cause many secondary effects (Roper and Reeves 2006).

Examination of the consequences of trisomy 21 has been complicated by the different chromosomal locations of mouse and human genes i.e. mouse equivalents of human chromosome 21 genes are scattered across 3 mouse chromosomes with 2/3 on chromosome 16 and the rest on chromosomes 10 and 17. The best mouse model is probably the Ts65Dn mouse which contains an extra chromosome 16 segment making it trisomic for about half of the human chromosome 21 genes (Kirsammer, *et al* 2008). The recent description of a mouse carrying an extra copy of human chromosome 21 has prompted much excitement – this model recapitulates many of the features of Down syndrome including craniofacial abnormalities, learning difficulties and heart defects (O'Doherty, *et al* 2005). No haematopoietic phenotype has been described to date, although studies are ongoing.

Unravelling the role of trisomy 21 has been attempted at cellular and molecular levels. Firstly, at a cellular level, two groups have looked at fetal liver haematopoiesis in autopsy material from Down syndrome fetuses (Chou, *et al* 2008, Tunstall-Pedoe, *et al* 2008). They both came to almost exactly the same conclusions by independent means. Namely that fetal liver multipotential progenitor cells (identified as CD34+CD38+ cells) show a striking increase in meg-erythroid potential with concomitant decrease in the CMP population. In addition these MEP populations were hyperproliferative with increased colony size,

increased clonogenicity on serial replating and increased erythroid and megakaryocytic output on transplantation to adult NOD/SCID mice. These features were restricted to fetal liver and not seen in bone marrow progenitors from the same fetuses (Tunstall-Pedoe, *et al* 2008). GATA-1 mutations were not seen in any of the samples (41 in total) – this contrasts with previous work showing GATA-1 mutations in 2/9 fetal liver samples (Taub, *et al* 2004). This may reflect the earlier gestational age of fetal samples used in this study (median 15 weeks (Tunstall-Pedoe, *et al* 2008) and range 13-23 weeks (Chou, *et al* 2008)). The absence of GATA-1 mutations allows the conclusion to be made that trisomy 21 leads to abnormalities in the haematopoietic progenitor compartment independent of GATA-1 mutations. This may provide an appropriate cellular substrate for GATA-1s expression to exert a phenotypic effect.

Another autopsy study revealed that peripheral blood samples from aborted Down syndrome fetuses have reduced CD34+ haematopoietic progenitors compared to age matched controls. This was also associated with reduced telomere length (Holmes, *et al* 2006). This contrasts with the above studies which showed normal CD34+ numbers in fetal liver and may reflect alteration in cellular localisation and transition between fetal liver and bone marrow in DS fetuses rather than an HSC deficiency *per se*.

Several different approaches have been taken to dissect out the molecular consequences of trisomy 21, and the molecular basis for co-operation between trisomy 21 genes and GATA-1s mutations. Thanks to the human genome project the complete sequence of chromosome 21 is known (Hattori, *et al* 2000) which encodes more than 350 genes. It is possible to look at these chromosome 21 genes and take a candidate gene approach for those that might play a key role in haematopoietic differentiation. Alternatively, since we are a long way from a complete description of the molecular basis for haematopoiesis, it is important not to discount genes that may have novel roles in this process – in order to identify such unknown genes expression profiling can be attempted.

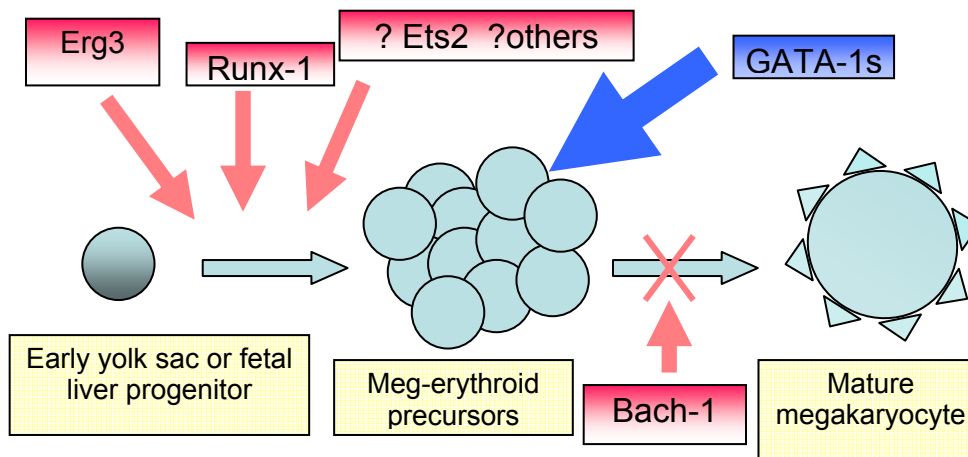
*A candidate gene approach:* Initial speculation pointed to a possible role for Runx-1 (AML-1) as the chromosome 21 co-operating gene. Many lines of evidence point to this gene being a prime candidate. Runx-1 is the commonest gene to be mutated in AML, it lies in the DS critical region, it is already known to exhibit gene dosage effects (albeit at the level of haploinsufficiency) with a predisposition to AML seen in familial platelet disorder (FPD) (Song, *et al* 1999), and it plays critical roles in erythroid and megakaryocytic differentiation. In addition it is known to physically interact with GATA-1 at both the N and C terminus (Elagib, *et al* 2003). Despite this, intensive study has been unable to prove

any association between Runx-1 perturbation and DS-AMKL. Importantly, Runx-1 overexpressing mice do not have an increased incidence of AMKL (Shimizu, *et al* 2008). Mutational analysis of Runx-1 in children with DS-AMKL has failed to identify any abnormalities (Rainis, *et al* 2003, Wechsler, *et al* 2002). A number of microarray studies have looked for overexpressed genes in DS-AMKL and have not seen any evidence of Runx-1 upregulation (Bourquin, *et al* 2006). Finally, the Ts65Dn mouse shows a myeloproliferative disease characterised by megakaryocytic hyperplasia and macrocytosis—using elegant crossing experiments between Ts65DN mice and AML-1 haplo-insufficient mice it is possible to restore disomy at the Runx-1 locus in Ts65Dn mice. This had no effect on the megakaryocytic phenotype (Kirsammer, *et al* 2008). This is probably the most compelling evidence to date that Runx-1 overexpression is not the dominant determinant of the leukaemic phenotype in Down syndrome. This does not necessarily mean it does not play a role, as extensively discussed above, transcription factors are dynamically regulated and both spatial and temporal patterns of expression are important as well as stoichiometry with other proteins in the cell. Runx-1 may well interact with other chromosome 21 genes to produce subtle changes in cellular potential.

Another candidate gene is *Erg3*. A recent study showed that it is expressed in HSCs, megakaryoblastic cell lines and AMKL samples (Rainis, *et al* 2005). In vitro differentiation of cell lines along the megakaryocytic lineage led to *Erg3* upregulation and forced expression of *Erg3* leads to a switch from erythroid to megakaryocytic gene expression in MEP like cell lines. *Erg3* binds directly to the SCL enhancer – a gene shown to be a key regulator of megakaryocytic development. Of interest another chromosome 21 gene *Ets-2* also binds to this region. This has led to a model of TAM/AMKL development, where combinatorial effects of Runx-1, *Erg3* and possibly *Ets2* lead to a skewing towards megakaryocytic fate and an expansion of megakaryocytic precursors in the fetal liver. This population is then vulnerable to effects of GATA-1 mutation leading to further proliferation and cellular transformation (Rainis, *et al* 2005). This is shown in figure 1.9.

*Gene expression studies:* A large study using samples from 24 DS-AMKL and 39 non-DS AMKL patients looked at gene expression differences between trisomy 21 associated TAM/AMKL and non-trisomy 21 cases. This showed that DS AMKL had a quite distinct gene expression profile (Bourquin, *et al* 2006). As expected a number of pro-proliferative genes initially identified in the mouse GATA-1s knock-in model (Li, *et al* 2005) were also overexpressed in DS-AMKL such as GATA-2, *myc* and *kit*. In addition increased expression of GATA-1 and two chromosome 21 genes *Bach-1* and *SON* were seen in DS-AMKL. *Bach-1* has been suggested to repress megakaryocyte differentiation by competing

with NF-E2 for essential cofactors needed to bind to megakaryocyte cis-regulatory elements (Toki, *et al* 2005). SON has homology with the myc family suggesting a possible role in cell cycle control. Bach-1 is added to the model proposed in fig 1.9 by providing a partial block to megakaryocytic differentiation in addition to the pro-proliferative effects of other chromosome 21 genes.



**Fig 1.9 Proposed model of combinatorial effects of chromosome 21 gene overexpression and additional mutations on development of AMKL.**

A second microarray pulled out 551 differentially expressed genes with 105 overexpressed in DS-AMKL and 446 in non DS-AMKL (Ge, *et al* 2006). This study confirmed the Bach-1 and GATA-1 upregulation reported previously. It also identified a gene - Bone marrow stromal-cell antigen 2 (BST2) - which was significantly underexpressed in DS-AMKL. This gene encodes a protein important in interaction between HSCs and stroma. The authors hypothesise that lack of BST2 leads to altered stromal interactions and increased sensitivity of DS blasts to chemotherapy (Ge, *et al* 2006).

Normal megakaryopoiesis is associated with micro-RNA down-regulation—presumably unblocking translation of genes responsible for megakaryocyte differentiation. In megakaryoblastic cell lines a number of miRNAs showed increased expression – which may reflect a differentiation block. Intriguingly, one of the upregulated miRNAs was encoded on chromosome 21 – miR 99a. Whether this contributes to the pathology of TAM/AMKL or the trisomy 21 MEP proliferation is unknown (Garzon, *et al* 2006).

### 1.3.2.5 Third genetic hit - mutations in Jak 2 and 3?

Since TAM often spontaneously resolves, the combination of trisomy 21 and GATA-1 mutations are insufficient to cause full-blown AMKL. A third genetic hit is needed. Overall, it seems likely that the third molecular event leading to full-blown AMKL will

differ between patients – this is consistent with most leukaemias in that a limited number of critical genes (often transcription factors), disrupted in a variety of ways, are common to many different leukaemias (Rosenbauer, *et al* 2005), but subsequent genetic events are more heterogeneous. Clinical evidence for this comes from a pair of identical twins sharing trisomy 21 and identical GATA-1 mutations. Both twins showed clonal evolution to AMKL but in one twin this was associated with a del(20q) and the other a trisomy 8 (Stark, *et al* 2002). This does not exclude a common third event followed by genomic instability but does show that cytogenetic abnormalities diverge at this point.

An initial attempt to identify the third hit used gene expression profiling to compare TAM and AMKL blasts from children with Down syndrome. These two groups showed remarkably similar gene expression profiles – consistent with their direct clonogenic relationship - but the investigators were able to pull out a few differentially expressed genes, most notably CDKN2C and N-myc as well as a tumour antigen PRAME (McElwaine, *et al* 2004). CDKN2C has already been identified as a mediator of GATA-1 cell cycle arrest. N-myc is a neuroblastoma associated oncogene overexpressed in TAM compared to AMKL. A second smaller study comparing AMKL and TAM did not pull out the same genes (Lightfoot, *et al* 2004). Instead a number of genes such as myosin light chain and N-acetyltransferase, not known to play roles in malignant transformation, were identified presumably reflecting altered cellular homeostasis in these cells.

In 2006 a group looking for aberrant tyrosine kinase activation in acute myeloid leukaemia identified a Jak3 activating mutation in an AMKL cell line – CMK (Walters, *et al* 2006). This approach involved a mass-spectroscopy based screen to identify tyrosine phosphorylated proteins which might be responsible for activation of key signal transduction pathways essential for the leukaemic phenotype. Further analysis showed the presence of an activating mutation in the pseudokinase domain of Jak-3 in these cells. Screening of Jak-3 in AMKL patients led to the identification of two further Jak-3 mutations - 1 in a patient with DS-AMKL and another in a non-DS AMKL sample. The authors went on to show that mice transplanted with BM cells containing a Jak3 activating mutation, develop a florid megakaryoblastic leukaemia. However, this does not appear to be secondarily transplantable suggesting, unsurprisingly, that Jak-3 activation alone cannot fully transform cells. This initial report was followed by several others – a Jak2 mutation was described in AMKL associated with trisomy 21 (although with no detectable GATA-1 mutation), again this was an activating mutation that produced megakaryocytic hyperplasia in a mouse model (Mercher, *et al* 2006). Further reports describe Jak-2 and Jak-3 activating mutations in AMKL (Hama, *et al* 2008, Kiyoi, *et al* 2007, Klusmann, *et al* 2007,

Norton, *et al* 2007, Sato, *et al* 2008) and also some loss of function mutations (De Vita, *et al* 2007). The association does not appear to be as tight as with GATA-1 – in fact Jak mutations are seen in TMD as well as AMKL (De Vita, *et al* 2008) and are not necessarily associated with GATA-1 mutations (Mercher, *et al* 2006) – in addition only a limited subset of DS-AMKL/TAM patients have any Jak mutations identified (Norton, *et al* 2007). It therefore appears likely that Jak mutations may be important determinants of disease progression and outcome but may not be the primary event associated with development of AMKL in most cases. It is interesting to note that these alterations in Jak activity would never have been identified by microarray studies as they result from point mutations leading to activation rather than changes in gene expression levels. The identification of key cellular events in malignancy requires multiple approaches and new developments in proteomics such as those used by Walters *et al* above are likely to provide crucial insights (Walters, *et al* 2006).

### **1.3.3 Chronic myeloid leukaemia**

In depth discussion of another haematological malignancy – chronic myeloid leukaemia is beyond the scope of this introduction. Study of this disease has provided many key insights into the molecular pathogenesis of malignancy (Melo and Barnes 2007). Its characteristic chromosomal translocation – t(9;22) bcr-abl, was the first consistent chromosomal abnormality to be associated with a specific human cancer and remains a paradigm for the role of chromosomal translocations in malignant transformation. It is mentioned here because of certain parallels with transient abnormal myelopoiesis, along with suggestive sequencing database information, which led to the hypothesis that GATA-1 could play a role in progression of CML from chronic phase to blast crisis.

#### **1.3.3.1 Molecular pathology**

CML is characterised by the presence of a reciprocal t(9;22)(q34;q11) translocation. This leads to the production of the bcr-abl oncoprotein with enhanced tyrosine kinase activity which leads to immortalisation of HSCs. CML is a bi- or tri-phasic disease. In the initial chronic phase patients display a massive expansion of the granulocytic lineage with increased circulating mature granulocytes and hepatosplenomegaly associated with expansion of granulocytic precursors in these organs. This chronic phase lasts for a median of 3-4 years before progressing to an acute (blast crisis) phase associated with a block to cellular differentiation and the appearance of blast cells. At this stage the disease behaves like an acute leukaemia and can have myeloid or lymphoid characteristics compatible with

transformation of a multipotent precursor/HSC. An intermediate phase, between chronic phase and blast crisis, associated with a progressive accumulation of immature precursors and blasts is also seen in some patients – known as accelerated phase. It is presumed that development of these acute phases is associated with acquisition of further mutations. According to the 2 hit theory of leukaemogenesis bcr-abl would provide a type I mutation affecting cellular proliferation and an additional type II mutation in a transcription factor should be necessary to provide the differentiation block. Despite a massive accumulation of knowledge as to the molecular consequences of the bcr-abl translocation, much less is known about this second event (Melo and Barnes 2007).

### **1.3.3.2 Similarities with transient abnormal myelopoiesis**

The blood film in chronic myeloid leukaemia typically shows a high white cell count, presence of variable numbers of immature precursors, often a marked basophilia and eosinophilia and thrombocytosis (Jaffe 2001). The erythrocyte lineage is usually unaffected. These appearances, especially the basophilia, share a resemblance to the blood film in TAM. In addition, in one study, 60% of patients progressing to myeloid blast crisis expressed megakaryocyte markers on their blast cells (Murakami, *et al* 1994). The reason for such a skewed granulocytic/platelet output with relative sparing of the erythroid lineage, despite bcr-abl being detectable in the stem cell compartment, has not been elucidated. Although GATA-1 mutations are said to be exclusive to DS-AMKL and TAM, the studies showing this specificity did not look at patients with CML (Wechsler, *et al* 2002). Therefore the mutational status of GATA-1 in CML is currently unknown.

### **1.3.3.3 The role of transcription factors in progression of disease**

The mechanisms of transformation to blast crisis appear to be varied and are not well understood. A number of potential mechanisms have been identified including differentiation arrest, genomic instability, telomere shortening and loss of tumour suppressor functions (Melo and Barnes 2007, Smith, *et al* 2003). There is accumulating evidence that acquired alterations in transcription factor activity do underlie transition to blast crisis in some patients (Zhang 2008). In common with AMKL this event is likely to vary between individuals and a single genetic cause for progression is extremely unlikely. Deletion or inactivation of p53, p16, Rb and Evi-1 have been reported but are demonstrable in <30% of cases of blast crisis (Pabst, *et al* 2006). Several different mutations in Runx-1 including translocations and point mutations have been described as have translocations involving HoxA9 but again these only represent sporadic cases (Zhang 2008). Despite this, their pathogenic role in development of blast crisis has been supported

by mouse transplantation assays showing that these translocations can act synergistically with bcr-abl in promoting acute leukaemia development in mice (Cuenco and Ren 2001, Dash, *et al* 2002). Gene expression profiling comparing blast crisis with chronic phase has identified further candidates including Runx-1 and Ets2 (Melo and Barnes 2007). In addition deregulation of the Wnt pathway - important for HSC interaction with the haematopoietic niche - was identified, as was up-regulation of PRAME – also identified in the TAM-AMKL transition (McElwaine, *et al* 2004, Melo and Barnes 2007).

In addition to direct effects of aberrant transcription factor expression, a number of other mechanisms may contribute to the development of blast crisis. In particular CML cells have been shown to accumulate mutations in the tumour suppressor genes p53 and Rb, they may also display suppression of apoptosis although this is controversial (Smith, *et al* 2003). Finally bcr-abl appears to promote DNA repair deficiency and therefore genomic instability. At least part of this is due to down regulation of the DNA repair enzyme, DNA-PK, via proteosome dependent degradation (Deutsch, *et al* 2001).

Although treatment of CML has been revolutionised by the availability of drugs targeting the abnormal tyrosine kinase fusion protein (such as Imatinib) (Druker, *et al* 1996), we are still a long way from complete eradication of the disease. These varied mechanisms of transformation to blast crisis, along with observations of the failure of Imatinib like compounds to eradicate bcr-abl containing stem cells (Elrick, *et al* 2005), suggest that molecularly targeted therapies will need to attack multiple pathways. However, it is clear from the above discussion that many leukaemias (and probably solid malignancies) share common themes when analysing the cellular and molecular basis for disease development and progression. Study of relatively rare malignancies that highlight specific features of these pathways can provide valuable insights into general mechanisms. Both Down syndrome AMKL and CML provide excellent biological models from which much can be learnt about the aetiology of malignancy and rational approaches to treatment.

## 1.4 Aims and Objectives

GATA-1 is an extensively studied haematopoietic transcription factor. The GATA-1s isoform is expressed at the protein level in haematopoietic cells and is clearly associated with human disease. Despite this, little is known about the normal biological role of GATA-1s. This thesis aims to address the distinct roles of GATA-1 isoforms in haematopoietic development and malignancy. A number of hypotheses have been formulated namely:



1. That the balance between GATA-1FL and GATA-1s isoform expression plays an important role in determining cell fate during haematopoietic development.
2. That GATA-1s exerts a distinct effect when expressed in cells carrying trisomy 21 compared to its effect in disomic cells.
3. That abnormalities of GATA-1 expression may play a role in chronic myeloid leukaemia, possibly by determining progression to accelerated phase or blast crisis.

In order to test these hypotheses a number of experimental approaches will be used:

1. A comprehensive analysis of the normal patterns of GATA-1 isoform expression in primary cells and cell lines, both in mature tissues and during embryonic development, alongside analysis of any possible additional isoforms.
2. Attempts to selectively knock-out isoform expression by generating targeted ES cells.
3. Analysis of the impact of isoform overexpression on the haematopoietic differentiation capacity of murine embryonic stem cells.
4. Analysis of the effects of over- and under-expression of GATA1 isoforms on cell fate, at the molecular and cellular level, in human haematopoietic cell lines with and without trisomy 21.
5. A mutational analysis of GATA-1 in CML blast crisis samples.

The results of these investigations are presented in the following chapters.

## 2 Materials and Methods

### 2.1 General solutions and consumables

The composition of chemical solutions, used in experiments detailed in the following chapters, is given in table 2.1 below. All laboratory chemicals were supplied by Sigma-Aldrich (Dorset, UK) unless otherwise stated.

Solution	Component	Concentration
<b>1 x PBS</b>	NaCl	137mM
	KCl	2.7mM
	Na <sub>2</sub> HPO <sub>4</sub>	10mM
	KH <sub>2</sub> PO <sub>4</sub>	1.76mM
<b>50 x TAE</b>	Tris (tris(hydroxymethyl)aminomethane) base	2M
	EDTA (Ethylene diamine tetra-acetate) disodium salt	50mM
	Glacial acetic acid	5.7% (v/v)
<b>1 x TE</b>	Tris (HCl) pH 8	10mM
	EDTA	1mM
<b>1x EB</b>	Tris (HCl) pH 8.5	10mM
<b>20 x SSC</b>	NaCl	3M
	Sodium citrate	300mM
<b>20 x MOPS</b>	MOPS (4-morpholine propane sulphonic acid)	1M
	Tris base	1M
	SDS (sodium dodecyl sulphate)	69.3mM
	EDTA	20.5mM
<b>Red cell lysis buffer</b>	NH <sub>4</sub> Cl	155mM
	KHCO <sub>3</sub>	10mM
	EDTA	0.1mM
<b>ES cell lysis buffer</b>	Tris (HCl) pH 8.5	100mM
	EDTA	5mM
	SDS	0.2%
	NaCl	200mM
	Proteinase K (Promega) (added just prior to use)	100µg/ml

<b>PBST (0.1%)</b>	PBS	1x
	Tween 20 (Polyoxyethylene sorbitan monolaurate)	0.1% (v/v)
<b>5% Milk PBST</b>	PBS	1x
	Tween 20	0.1% (v/v)
	Marvel skimmed milk powder	5% (w/v)
<b>Southern blot alkaline transfer buffer</b>	NaOH	0.4N
	NaCl	1M
<b>Southern blot neutralisation II buffer</b>	Tris-Cl (pH 7.2)	0.5M
	NaCl	1M
<b>Southern blot high stringency wash buffer</b>	SDS	0.1%
	SSC	0.1x
<b>Southern blot low stringency wash buffer</b>	SDS	0.1%
	SSC	2x
<b>Nuclear extract buffer A</b>	HEPES (4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid) pH 7.9	10mM
	MgCl <sub>2</sub>	1.5mM
	KCl	10mM
	Just before using add:	
	PMSF (phenylmethylsulphonyl fluoride)	0.2mM
	DTT (Dithiothreitol)	0.5mM
<b>Nuclear extract buffer C</b>	HEPES pH 7.9	20mM
	Glycerol	25% (v/v)
	NaCl	420mM
	MgCl <sub>2</sub>	1.5mM
	EDTA	0.2mM
	Just before using add:	
	PMSF	0.2mM
	DTT	0.5mM
<b>Western blot stripping buffer</b>	SDS	2% (w/v)
	Tris (pH 6.8)	50mM
	β-mercaptoethanol	100mM

<b>Propidium iodide mix</b>	Propidium iodide	40µg/ml
	RNase (DNase free)	100µg/ml
	PBS	1x
<b>Haemin (4mM stock)</b>	Haemin (C <sub>34</sub> H <sub>32</sub> N <sub>4</sub> O <sub>4</sub> FeCl)	50mg
	Dissolve in:	
	NaOH 0.5M	3.3ml
	Tris (pH 7.8) 0.1M	0.99ml
	ddH <sub>2</sub> O	14.9ml
	filter sterilise prior to use	
<b>Plasmid lysis buffer</b>	EDTA	5mM
	Sucrose	10% (w/v)
	SDS	0.2% (w/v)
	NaOH	100mM
	KCl	60mM
	Bromophenol blue	0.005% (w/v)
<b>10 x DNA loading buffer</b>	Bromophenol blue	0.25% (w/v)
	Glycerol	30% (v/v)
<b>6 x sucrose gel loading buffer</b>	Bromophenol blue	25mg
	Xylene cyanol	25mg
	Sucrose	4g
	ddH <sub>2</sub> O	10ml

**Table 2.1: Solution composition**

General plastic lab-ware and other consumables were obtained from a variety of suppliers listed below. Manufacturers and suppliers of equipment, reagents and commercial kits are detailed in the main body of the text in parentheses after the equipment / reagent / kit name.

**ABgene (Vancouver, Canada)**

Thin walled 0.2ml and 0.5ml PCR tubes

**Applied Biosystems (Warrington, UK)**

MicroAmp 96-well optical reaction plates and adhesive covers

**BDH laboratory supplies (Poole, UK)**

1mm thick twin-frosted microscope slides

**Becton-Dickinson Labware (Le Pont de Claix, France)**

BD Falcon <sup>TM</sup> conical tubes 15ml and 50ml

BD Plastipak<sup>TM</sup> syringes (1ml, 2ml, 5ml, 10ml, 20ml and 50ml)

Polypropylene 2059 tubes

Polystyrene 2054 5ml round-bottomed FACS tubes

**Bibby Sterlin Ltd. (Bargoed, UK)**

Petri dishes 100mm diameter

Universal containers

**Corning Inc (Poole, UK)**

Tissue culture plates – 6-well, 12-well, 24-well and 96-well

**Dow Chemical Co (Midland USA)**

Saran-wrap

**Eppendorf (Hamburg, Germany)**

1.5ml microfuge tubes

**Greiner Bio-One (N.Carolina, USA)**

Tissue culture flasks with filter cap - 25cm<sup>2</sup>, 75cm<sup>2</sup> and 175cm<sup>2</sup>

**Nunc International (Thermo-Fisher Scientific, Roskilde, Denmark)**

1.5ml cryo-tubes

Nalgene filter unit 500ml

**Swann-Morton Ltd (Sheffield, UK)**

Disposable Scalpels

**Whatmann International Ltd (Maidstone, UK)**

3MM filter paper

## 2.2 Media

The following base media were used for tissue culture:

Base media	Abbreviation	Supplier
Dulbecco's Modified Eagles medium	DMEM	Invitrogen, Paisley, UK
Knockout Dulbecco's Modified Eagles medium	KO-DMEM	Invitrogen
Iscove's Modified Dulbecco's medium	IMDM	Invitrogen
RPMI-1640 medium	RPMI	Invitrogen
Leibovitz's L-15 medium	L15	Invitrogen

Table 2.2 below lists the cell lines used along with their basic characteristics and culture requirements:

Cell line	Adherent or suspension	Culture medium	Special considerations	Seeding density
K562	Suspension	<u>Complete RPMI</u>		Jurkat,
Jurkat		RPMI 1x		P815,
WEHI-3B		FCS 10% (v/v)		U937:
P815		L-glutamine 1x (2mM)		3x10 <sup>5</sup> /ml
U937		Penicillin and Streptomycin		WEHI-3B
Raji		1x (final concentration 50µg/ml each)		1x10 <sup>5</sup> /ml
Meg01	Suspension (semi-adherent)	Complete RPMI –as above	Dislodge adherent cells by pipetting	Raji 4x10 <sup>5</sup> /ml

E14 (murine ES cells)	Adherent	<u>Complete Knockout DMEM</u> KO DMEM 1x ES cell FCS 15% (v/v) L-glutamine 1x PenStrep 1x B-mercaptoethanol 0.1mM LIF 1000u/ml Non-essential amino acids 0.1mM Gentamicin 50µg/ml	See section 2.14	Split every 2 days 1:3 to 1:6
HEK293, COS7, NIH3T3, RAW264	Adherent	<u>Complete DMEM</u> DMEM 1x FCS 10% (v/v) L-Glutamine 1x PenStrep 1x	RAW264 semi- adherent remove mechanically, do not allow to become confluent	2-4x10 <sup>5</sup> /ml
M07E	Suspension	RPMI 1x FCS 20% (v/v) L-glutamine 1x PenStrep 1x GM-CSF 10ng/ml	Split 1:2 to 1:3 every 2 days	4x10 <sup>5</sup> /ml
32D B6SutA	Suspension	complete RPMI (see above) plus 10% (v/v) WEHI-3B conditioned media	WeHI 3B conditioned media used as a source of IL-3	32D: 5x10 <sup>5</sup> /ml  Split 1:10- 1:50 every 2-3 days
P388	Semi- adherent	<u>complete L15</u> L15 1x FCS 10% (v/v) PenStrep 1x	Detach by tapping sharply	3x10 <sup>5</sup> /ml
FDCPmix	Suspension	Iscoves MDM 80% FCS 20% muIL-3 10ng/ml (or 10-20% WeHI-3B conditioned media)	Low cell recovery – thaw into 24 well plates and add fresh IL-3 after 8 hours, then daily until established	1x10 <sup>6</sup> /ml split 1:2 every 3-4 days

Table 2.2 Cell lines and culture requirements

## 2.3 Bacterial culture media

Medium	Component	Concentration
Luria-Bertani (LB) Broth	Tryptone	1% (w/v)
	Yeast extract	0.5% (w/v)
	NaCl	1% (w/v)
	Sterilise by autoclaving	
LB agar plates	Agar	1.5% (w/v)
	LB-broth	1 x
	On cooling add:	
	Ampicillin	50µg/ml
	OR	
	Kanamycin	50µg/ml

**Table 2.3: Bacterial culture media**

## 2.4 Primers, RNAi reagents and antibodies

All primers were synthesised by VH Bio (Gateshead, UK) using reverse phase cartridge purification unless otherwise stated. All sequences are reproduced in 5' to 3' orientation.

### 2.4.1 *GATA-1* primers

Primer name	Species	Forward sequence	Reverse sequence	Annealing temp °C
GATA-1 (exon 4/5)	Mouse	cgaatgattgtcagcaaacg	accagctaccacatgaagc	55
Alt spl 1	Mouse	atcctctgcatcaacaagcc	tccatactgttgagcagtgg	52
Alt splice version 2	Mouse	actcgtcataccactaaggtg	aagggtgtccaagaaggtgttg	55
Exon1/3 splice	Mouse	tcaacaagcccagctctttcag	tcaaggctattctgtgtacc	55
Exon1/3 splice site 4bpmm	Mouse	tcaacaagtagtggtctttcag	tcaaggctattctgtgtacc	55
Human alt spl	Human	ctccgcaaccaccagcccag	tatggtgagccccctgggac	58
Exon 1 & 6	Human	ctccgcaaccaccagcccag	tcatgagctgagcggagccac	55
Exon 3 & 6	Human	ccgtggaagatctggatgg	tcatgagctgagcggagccac	55

**Table 2.4: *GATA-1* primers**



### 2.4.2 Primers used for vector construction

Name	Species	Forward primer sequence	Reverse primer sequence	Annealing temp °C
BFT4 FL GATA	mouse	gttcccatggatttctg	gctattctgtgtaccttcaagaac	52
BFT4 GATA1s	mouse	caacagtatggagggaattcc	gctattctgtgtaccttcaagaac	55
GATA-1FL	human	accatggagttccctggcctg	tcatgagctgagcggagccac	58
IMAGE	human	ctgggatcacactgagcttg	ttttatcaggactttggtgg	52

**Table 2.5: Primers used for vector construction**

### 2.4.3 Primers for site-directed mutagenesis

Name	Forward primer	Reverse primer
GATA-1s SDM	gccccactcagtctctgaaggtaacacag	ctgtgtaccttcaagagactgagtggggc
GATA-1FL SDM	ccactgctcaacagtacagagggaattcctgg	ccaggaattccctctgtactgttgagcagtgg
GATA-1s targeting	accccagtgttcccatcgatttctctggtc	gaccaggaaaatcgatgggaacactgggggt
GATA-1FL targeting	tatccactgctcaacagtctagagggaattcctgggggc	gccccaggaattccctctagactgttgag cagtggata
GATA-1FL His tag	gccccactcagttctgaaggtaacacag	ctgtgtaccttcagaactgagtgggggc

**Table 2.6: Primers for site directed mutagenesis**

### 2.4.4 Primers for RT-PCR transcriptional profiling

Name	Species	Forward primer	Reverse primer	°C
β-actin	Mouse / human	tgaaccctaaggccaaccgtg	gctcatagctcttctccagg	58
GAPDH	human	caaggctgagaacgggaag	ggtggtgaagacgccagt	55
Glycoprotein 1b	human	tgcttttgcctttctccac	ccccttctctgggtatttg	52
CD41	human	ggtgagagggagcagaacag	accttgagagggttgacagg	55
Cyclin D2	human	ttacctggaccgttcttgg	ctgaggcttgatggagtgtgc	55
Platelet factor 4	human	ctgttctctgggggtgctg	gctatcagttgggcagtgg	55

$\beta$ -globin major	human	ggctcacctggacaacctc	tggacttagggaacaaaggaac	55
Glycophorin A	human	tagttctggaggctgggaag	tctgatgaaaggatgtgtttgg	52
$\epsilon$ -globin	human	gcaagatgaatgtggaagagg	caggatggcagagggagac	55

**Table 2.7: Primers for RT-PCR transcriptional profiling**

### 2.4.5 Primers for CML mutational analysis

#### 2.4.5.1 Direct sequencing:

Names	Forward primer	Reverse primer
HuGP1/4 (ex2)	tgaggtgatggagtgggaggagg	ggtcggcacatccattgagaagc
HuGP5/8 (ex3)	ccatgttgggggtgctgggaacca	gtgggggtggagaggagaagaggga
HuGP9/10B (ex4)	aaaaaggacagggaagttgaggtg	gcagtagcttctgtaatcatgag
HuGP11/12 (ex5)	cctcacttcttgggtcctcctga	agagaggtagaacaggaacagagt
HuGP13/14 (ex6A)	gcacccaaaaattatcttaccctg	tgaagccacctccccacaattccc
HuGP15/16 (ex6B)	ctggaaaagggaagaaacggg	aaggttcaagccagggggtgcctg

**Table 2.8: CML direct sequencing primers**

#### 2.4.5.2 WAVE

Primer name	Forward	Reverse	T <sub>m</sub>	Buffer	Elution temps
2C1F / 2C4R	aaaggagggaagaggagcag	aagcttcagccatttctga	60	A	60.2 / 61.2 / 62.2
3.1F / 3.1R	ggaacttgccaccatgttgg	agccgctctgtcttcaaagtctc	58	A	61.9 / 62.9
3.2F / 3.2R	ggatcccagggggtcaccata	gtgggggtggagaggagaagaggga	58	B	60.4 / 61.1 / 62.1
4F/4R	gacagggaagttgaggtgggag	cttctgtaatcatgagaacagcg	56	B	58.7 / 60.7 / 62.7
5F/5R	ggcatcacctgtaaacaagcc	gggcagtgtggcatgaagac	58	B	62.3 / 64.3
6.1F / 6.1R	gtgtccctggttgacacagag	ggcccagtgtcaggcctgaagcca	58	B	62.6 / 63.6
6.2F / 6.2R	gaaggaccagctggtggctttatg	ggctacaagaggagaaggac	60	B	63.9 / 64.9

**Table 2.9: CML WAVE primers**

### 2.4.6 Primers and probes for ES cell targeting

Name	Forward primer	Reverse primer	Tm
Exon 2 xho1/kpn1 # 1	attctcgagatagacatggacggccaa ag	aatggtaccaagaggctgcaggt caagtg	53
Exon 3-6 EcoRV	attgatatcacttgacctgcagcctcttg	gaggggaagcttgaactgcac	55
Exon 2 xho1/kpn1 # 2	attctcgagcaaagaaggcagtgaggac ac	aatggtaccacagcttcgtacaca cattccag	55
PCR+ screening plasmid	gccgagctgtgtgtagtaagg	attgcatcgcatgtctgag	52
PCR screen # 1	gccgagctgtgtgtagtaagg	ctaccggtggatgtggaatg	55
Nested ext primers	ggaaagaggggaaaggtgag	cagaaagcgaaggagcaaag	55
Nested int primers	gccgagctgtgtgtagtaagg	ctaccggtggatgtggaatg	55
S.blot BamH1 probe	tatcccaaattctggccttg	ggtcagagttcctggagcag	52
S. blot AflII probe	agtagagccatcagcccatc	gtttctctaatacttcccatcttc	55

**Table 2.10: ES cell gene targeting primers**

### 2.4.7 Probe/primer pairs for real-time PCR

Target	Sequence	Component
GATA-1FL	ccagggttaatccccagagg	5' primer
	cccagggttaatccccagag	3' primer
	ccagaggctccatggagttccct	Probe (FAM/TAMRA)
GATA-1s	accaccagcccactct	5' primer
	cagttgaggcagggtagagc	3' primer
	ccatatgccggctgggccta	Probe (FAM/TAMRA)

**Table 2.11: qPCR primer/probe sets**

### 2.4.8 RNAi duplexes

Duplex	Sense sequence	Antisense sequence	Exon targeted
HuGATA duplex 5	ggacaggccacuaccuauuu	cauagguaguggccuguccuu	4
HuGATA duplex 6	acgcugaggccuacagacauu	ugucuguaggccucageguuu	2
HuGATA duplex 7	gcugguggcuuuauugguguu	ccaccauaaagccaccagcuu	6
HuGATA duplex 8	ccaagaagcgccugauuuguuu	acaauccaggcgcuucuugguu	4

**Table 2.12: RNAi duplexes**

## 2.4.9 Antibodies

### 2.4.9.1 For Western blotting

Name	Raised in / immunogen	Secondary antibody	Species reactivity of primary antibody	Working dilution 1°/ 2°	Suppliers
GATA-1 C-terminus (M20)	Goat anti-mouse polyclonal	Protein G HRP (bacterial protein)	human, mouse	1:100 / 1:5000	SantaCruz Biotech / Bio-Rad
GATA-1 N-terminus (N6)	Rat anti-mouse monoclonal	Goat anti-rat IgG HRP	mouse	1:1000 / 1:1000	SantaCruz Biotech.
Actin (JLA20)	Mouse anti-chicken monoclonal	Goat anti-mouse IgM HRP	human	1:10000/ 1:2000	Calbiochem
GAPDH	Rabbit anti-human polyclonal	Goat anti-rabbit HRP	human, mouse	1:5000/ 1:2000	Abcam
His(C-terminal)	Mouse monoclonal	Goat anti-mouse IgG	n/a	1:5000 / 1:10000	Invitrogen

**Table 2.13: Antibodies for Western blotting**

HRP = Horseradish peroxidase

### 2.4.9.2 Antibodies for flow cytometry

Name	Species	Fluorochrome	Channel	Supplier
Glycoprotein Ib (CD41a)	Mouse anti-human	APC	FL4	BD Pharmingen
Glycophorin A (CD235a)	Mouse anti-human	PE-Cy5	FL3	BD Pharmingen
FcR blocking reagent	Anti-human	-	-	Miltenyi Biotec

**Table 2.14: Antibodies for flow cytometry**

## 2.5 Cell Culture Methods

Table 2.2 lists the various cell lines utilised in this thesis along with their routine culture conditions and any special notes. Methods for the thawing, maintenance and freezing down of cell lines are given below. Biological assays involving the use of cell lines are described in section 2.13, embryonic stem cell protocols are described in section 2.14. All procedures involving cell culture were carried out using sterile precautions i.e. all manipulations were performed in a laminar flow hood with HEPA filtration along with use of pre-sterilised pipettes, glassware and plasticware, 70% Ethanol spray on all surfaces and non-autoclavable equipment and use of a “no-touch” technique. All tissue culture reagents were sterilised by autoclaving or passage through a 0.22µm filter. Unless otherwise mentioned, all centrifugation steps for cell culture were performed at 1200rpm for 5 minutes using a Centaur 2 bench-top centrifuge (MSE, London). All cultures were incubated at 37°C / 5% CO<sub>2</sub>/ 95% humidity.

### 2.5.1 Thawing of cell lines from frozen stocks

Cell lines were grown up from stocks stored in liquid nitrogen (-196°C). Recovery of cells from liquid nitrogen was performed by rapid thawing in a 37°C water bath. Thawed cells were washed in 5ml of the appropriate culture medium (warmed to 37°C), followed by centrifugation (1200rpm, 5 minutes). The supernatant was then discarded and the cell pellet resuspended in 5ml of fresh warmed culture medium before transfer to 25cm<sup>2</sup> tissue culture flasks.

### 2.5.2 Cell counting

Cells were counted using a modified Neubauer Haemocytometer. 10µl of cell suspension was loaded into the counting chamber and the number of cells in the 4 x 4 grid was counted. This number was then multiplied by 10<sup>4</sup> to give the number of cells per ml of suspension. In cases where accurate counting of live cells was required, dead cells were excluded using Trypan blue: 2µl of Trypan blue was added to 8µl of cell suspension, incubated at room temperature for 5-15 minutes and then the sample was loaded into the haemocytometer chamber as before. Dead cells were stained blue and were excluded from the cell count.

### **2.5.3 Maintenance of cell lines in culture**

Cell lines were routinely passaged when they reached 70-80% confluency or according to cell line specific requirements as detailed in Table 2.2. Methods differed depending on whether cells grew as adherent or suspension cultures. In order to minimise acquisition of new mutations, low passage cells were used whenever possible. Cells were routinely cultured in the presence of antibiotics (Penicillin and Streptomycin) unless this was contra-indicated (e.g. in the 48 hours prior to transfection with some transfection agents – see section 2.12.3 below). Cultures were visually inspected daily under an inverted microscope to assess confluency and rule out infection. In addition ES cell cultures were inspected for evidence of differentiation as described in section 2.14.1.

#### **2.5.3.1 Passage of adherent cell cultures**

Media were aspirated and adherent cells were washed with 5-10 ml of warmed phosphate buffered saline (PBS) to remove any remaining traces of fetal calf serum. The PBS was replaced with 0.05% Trypsin (w/v) (Invitrogen) (1ml for 25cm<sup>2</sup> flask, 3ml for 75cm<sup>2</sup> and 5ml for 175cm<sup>2</sup>) warmed to 37°C. After incubation for 2-5 minutes at 37°C the cells were observed under an inverted microscope to ensure adequate detachment. Trypsinisation was stopped by addition of 3-7ml of warmed media and the cells were resuspended by vigorous pipetting. The resulting suspension was centrifuged at 1200rpm for 5 minutes. After removal of the supernatant cells were resuspended in an appropriate volume of fresh culture media and transferred to a new flask. In general cells were split between 1:3 and 1:10. In the case of new cell lines or cell lines with particularly fastidious requirements the cell suspension was counted and then seeded at the published optimum density (listed in Table 2.2).

#### **2.5.3.2 Passage of suspension cell cultures**

Suspension cultures were centrifuged at 1200 rpm for 5 minutes, resuspended in 3-10ml of fresh culture medium and a proportion of this was transferred to a sterile tissue culture flask depending on the split required, as detailed for adherent cells above.

### **2.5.4 Freezing down of cell lines**

Frozen stocks of early passage cells were established for all cell lines. Cells for freezing were washed in PBS, spun down and resuspended at a concentration of approximately

$1 \times 10^7$ /ml in a mixture of 90% culture media, 10% v/v dimethylsulphoxide (DMSO) to prevent formation of ice crystals. 1ml aliquots were transferred to 1.5ml cryovials and placed in a freezing container (Nalgene, Hereford) containing room temperature isopropanol. The container was placed at  $-80^{\circ}\text{C}$  overnight to allow gradual cooling ( $1^{\circ}\text{C}$  per minute), then transferred to liquid nitrogen tanks ( $-196^{\circ}\text{C}$ ) the following day.

## **2.6 Animal methods**

### ***2.6.1 Isolation of mouse bone marrow***

C57/BL6 male mice were sacrificed humanely by  $\text{CO}_2$  asphyxiation. Femurs were dissected by cutting just below the knee joint and above the hip joint and placed immediately in ice-cold PBS. Bones were separated from surrounding soft-tissue, cut at both ends and repeatedly flushed through with 1ml of ice-cold PBS via a 26 gauge needle. 6mls of red cell lysis buffer was added to the resultant cell suspension, the mixture incubated at room temperature for 5 minutes followed by centrifugation for 10 minutes at 1500rpm. The cell pellet was resuspended in 10ml PBS and cells counted using a modified Neubauer haemocytometer (see 2.5.2 above). Cells were then pelleted once more and resuspended in an appropriate volume of Trizol reagent (see 2.8.1.1 below) for subsequent RNA extraction.

### ***2.6.2 Mouse embryonic dissections***

Following timed mating, C57/BL6 female mice were sacrificed humanely by cervical dislocation on days E8.5, E9, E10.5 and E12.5. The uterus was dissected out using forceps and placed in ice-cold PBS. All subsequent steps were performed using micro-dissection instruments under a dissecting microscope following published protocols (Nagy 2003). Initially individual embryos were dissected away from the muscular wall of the uterus, Reichert's membrane and the visceral yolk-sac. Embryo gestation was confirmed by morphological features and somite counting. For day 8.5 and 9 embryos the embryo and yolk sac were separated and placed in individual wells of a 12-well tissue culture plate containing ice cold PBS. At day 10.5 the AGM region was dissected away from the rest of the fetal body and placed in PBS, fetal blood was collected by placing the dissected embryo into 1ml of ice cold PBS and allowing the blood to flow into the solution. At day 12.5 the embryo was rapidly and humanely sacrificed by severance of the fetal head from the fetal body. The fetal liver was easily recognisable and was micro-dissected free from

the rest of the fetal body. All samples were washed in ice-cold PBS, centrifuged at 3000rpm for 4 minutes and the resuspended in 50µl PBS and 50µl 2x Laemmli sample buffer (Sigma-Aldrich) before being stored at -20°C. Prior to use in western blotting whole cell lysates were DNase treated to reduce sample viscosity as described in section 2.8.2.

## **2.7 Molecular Biology Methods: DNA**

### ***2.7.1 Preparation of genomic DNA***

The following protocol was used to extract genomic DNA from ES cell colonies growing in 24 well tissue culture plates:

---

Add 100µg/ml Proteinase K (Promega) to ES cell lysis buffer prior to use  
Aspirate medium from cells, wash twice in 1ml cold PBS  
Add 250µl ES cell lysis buffer (with proteinase K) to each well  
Seal sides of plate with tape and place in humidified container, incubate 55°C overnight  
The following day transfer lysates to a fresh 1.5ml microcentrifuge tube  
Add 0.7 volumes (175µl) isopropanol and mix by inversion until a precipitate is visible  
Spin 5 minutes at 5000rpm  
Pour off supernatant and wash pellet in 0.5ml 70% Ethanol  
Spin 5 minute 500rpm, remove supernatant and briefly air-dry pellet  
Resuspend in 100µl TE, heat 65°C 5 minutes to evaporate residual ethanol and shake at 4°C overnight prior to use

---

This buffer and protocol could also be used for extraction of DNA from standard cell lines such as HEK293.

### ***2.7.2 Purification, concentration and quantification of DNA***

DNA was purified for use in downstream applications by use of phenol/chloroform extraction followed by ethanol precipitation or the use of commercial PCR/gel purification kits as described in section 2.7.5 below. DNA was concentrated by ethanol precipitation followed by resuspension in a reduced volume of TE or ddH<sub>2</sub>O.



---

**Phenol chloroform extraction**

---

Add equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 v/v/v) to DNA solution

Mix by vortexing

Centrifuge 14,000rpm for 5 minutes

Transfer upper aqueous (DNA) phase to a fresh microcentrifuge tube and perform ethanol precipitation.

---

---

**Ethanol Precipitation**

---

Add 1/10<sup>th</sup> volume 3M sodium acetate (0.3M final concentration)

Mix well then add 2 volumes ice-cold 96-100% ethanol

Incubate on ice 15-30minutes

Centrifuge 14,000rpm for 10 minutes at 4°C

Carefully remove supernatant then add 750µl 70% ethanol, centrifuge 14,000rpm for 2 minutes at 4°C

Remove supernatant, air dry pellet for 15 minutes (in sterile hood if needed)

Redissolve in appropriate volume of Tris-EDTA (TE) buffer or ddH<sub>2</sub>O

Heat 2-3minutes at 45°C to evaporate residual ethanol

---

DNA was quantified by measurement of the amount of UV absorbance at 260nm. 3µl of the test sample was diluted with 297µl of TE buffer and the OD<sub>260</sub> measured using a Biophotometer (Eppendorf). The OD<sub>260</sub> was automatically multiplied by the dilution factor and the extinction co-efficient to give a concentration in µg/ml. Purity was assessed by measurement of the OD<sub>260</sub>:OD<sub>280</sub> ratio aiming for levels between 1.7 and 1.9. Levels below 1.7 indicate contamination by proteins or organic chemicals. Quantitation was also performed in some cases by gel electrophoresis with comparison against standard concentrations of DNA used in DNA marker ladders, this method also established contamination with other DNA (additional bands) and RNA as well as revealing plasmid degradation (lack of supercoiled fast migrating band).

### **2.7.3 Restriction enzyme digests**

Restriction enzymes were obtained from a variety of commercial suppliers (New England Biolabs, Roche, Fermentas, Invitrogen). A standard diagnostic digest comprised:

---

**Restriction enzyme digest**


---

DNA sample (eg plasmid miniprep)	5µl (approximately 0.5µg)
Restriction enzyme (RE) buffer (10x)	1µl
ddH <sub>2</sub> O	3.5µl
Restriction enzyme (10 units/µl)	0.5µl (5 units)
Mix gently, incubate at 37°C (or alternative temperature as specified in datasheet) for 1-2 hours.	

---

Where possible double digests were performed using compatible buffers as determined by compatibility charts ([www.neb.com](http://www.neb.com)). If not, then sequential digests were performed with removal of the first buffer by use of a Qiagen PCR purification kit (section 2.7.5) prior to the second round of digestion.

For diagnostic digests, results were analysed by gel electrophoresis (see 2.7.4 below). The test sample was run alongside uncut vector DNA to allow distinction between bands due to supercoiled and nicked plasmid DNA and linear fragments.

See section 2.7.7.1 for restriction enzyme digestion of high molecular weight genomic DNA for use in Southern blotting, and section 2.12.1.2 for restriction enzyme digests prior to DNA ligation.

### **2.7.4 Gel electrophoresis of DNA**

DNA molecules were size separated on 0.8-2% agarose gels depending on the expected molecular weight of the products:

Size of DNA fragments	% agarose w/v dissolved in Tris-acetate - EDTA buffer (TAE)
100bp-2kb	2%
200bp-3kb	1.5%
500bp-7kb	1.0%
800bp-10kb	0.7%

---

Electrophoretic grade agarose was added to TAE buffer and dissolved by microwaving. After partial cooling, ethidium bromide was added (0.5µg/ml) to allow DNA visualisation under UV light. The gel was cast in a tray with appropriate sample combs. Once set, the

gel was placed in a suitable horizontal electrophoresis tank, combs removed and TAE buffer was added to the tank until just covering the surface of the gel. Samples were mixed with 1/10<sup>th</sup> volume of 10x loading buffer. 5µl of a marker either (i) 1kb ladder (Invitrogen) or (ii) Hyperladder I (Bioline) (for fragments 1-12Kb) or (iii) Hyperladder IV (Bioline) (for fragments 100bp-1kB), were loaded alongside to allow DNA sizing and/or quantitation. Electrophoresis was performed at 90-120V depending on the size of the electrophoresis tank (max 5V/cm) until the band of visible loading dye reached approximately 3cm from the end of the gel. The gel was then carefully removed and visualised under UV illumination at the 365nm wavelength (using suitable protective apparatus). Photographs were taken using an Alpha Imager (Alpha Innotech, San Leandro, USA) gel imager system.

### ***2.7.5 Gel extraction and PCR purification***

If DNA was to be used for downstream applications the gel was exposed to the minimum duration of UV light at reduced wavelength (302nm) to avoid DNA damage. Bands were excised using a clean scalpel blade and placed in a microcentrifuge tube. DNA extraction was performed by utilisation of spin-column technology exploiting the ability of DNA to adsorb to a silica membrane, allowing removal of impurities by wash buffers followed by elution of the DNA with Tris buffer or water. This was achieved by use of a QIAquick Gel extraction kit (Qiagen) according to manufacturer's instructions. Similar technology was also used to directly purify PCR products using the QIAquick PCR purification kit (Qiagen), again procedures were carried out according to manufacturer's instructions.

### ***2.7.6 Sequencing of DNA***

Sequencing was performed by a variety of sources (i) Technology services department – Beatson Institute, Glasgow, (ii) Agowa commercial DNA sequencing service ([www.Agowa.de](http://www.Agowa.de)) (iii) Weatherall Institute of Molecular Medicine (WIMM) sequencing service, Oxford. All services employed Sanger sequencing methodology based on Dye terminator cycle sequencing technology. The initial labelling reaction was performed by the sequencing service in (i) and (ii) above. Samples sequenced post WAVE analysis (iii) were labelled using a BigDye terminator sequencing kit v3.1 (Applied Biosystems) in the following 20µl reaction carried out in 96 well thin-walled PCR plates:

---

Purified PCR product	4µl
Big Dye ready reaction pre-mix	4µl
Big Dye Dilution buffer (5x)	4µl
Primer (2µM)	1µl
Ultrapure water	7µl

---

Mix well, spin briefly then overlay with mineral oil and run on a thermal cycler overnight according to manufacturer's instructions, the next morning ethanol precipitate products as described in section 2.7.2

---

Cycle sequencing products were then analysed on an ABI 3100 capillary array sequencer (Perkin Elmer, Beaconsfield, UK).

### **2.7.7 Southern blotting**

#### **2.7.7.1 DNA digestion and gel electrophoresis**

To ensure complete digestion of high-molecular weight genomic DNA a 50µl reaction was performed with 2µg genomic DNA, 5µl 10x buffer, 0.5µl Bovine serum albumin (BSA) (100x), 3µl (30units) of restriction enzyme and xµl ddH<sub>2</sub>O to a final volume of 50µl. The reaction was stored at 4°C for several hours prior to the addition of enzyme and kept on ice for a couple of minutes after addition of enzyme. A second 3µl aliquot of enzyme was added after 12 hours followed by a further 24 hour incubation. If stored at 4°C prior to analysis then the digest was heated to 56°C for 3 minutes to disrupt any base pairing between cohesive termini. 7.5µl of 6x sucrose gel loading buffer were added and the 50µl of this digested DNA was loaded into wells of a 0.7% (w/v) agarose gel cast in 1xTAE buffer with 0.1µg/ml ethidium bromide. 10µl of Hyperladder I were loaded in the first and last lanes to allow size discrimination. Following application of a 110V current for 10 minutes to allow the samples to run into the gel, electrophoresis was then performed at low voltage (25V) overnight. The gel was then photographed alongside a UV ruler to allow subsequent sizing of bands.

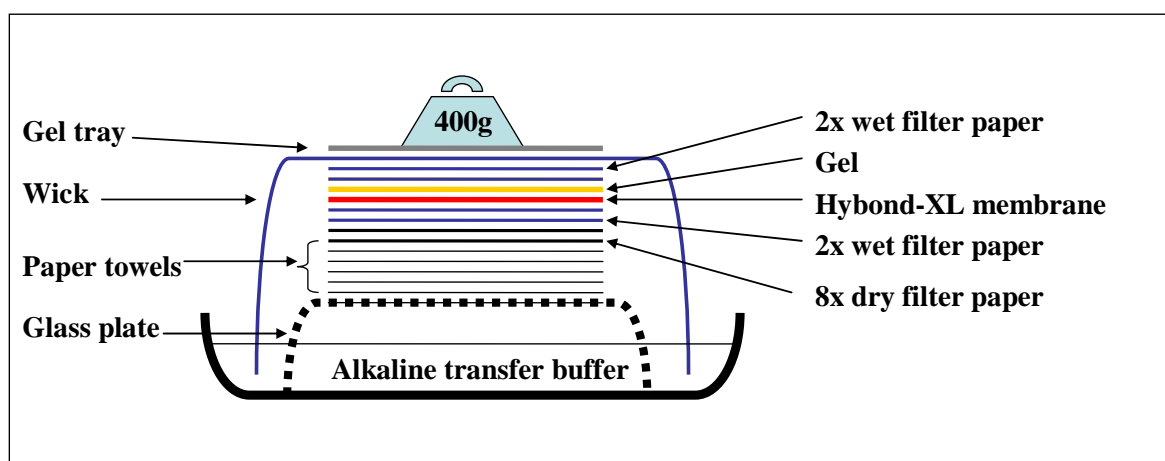
#### **2.7.7.2 Preparation of the gel and membrane for transfer**

The gel was trimmed, marked in the bottom left-hand corner and transferred to a glass dish. DNA denaturation was performed by soaking the gel at room temperature in several volumes of alkaline transfer buffer with constant gentle agitation on a rotary platform.

After 15 minutes the solution was exchanged for fresh alkaline transfer buffer and agitation continued for a further 20 minutes. A nylon Hybond-XL (Amersham International/GE Healthcare) membrane was used for nucleic acid transfer, this was cut to the same size as the gel along with 2 sheets of thick filter paper (Whatman 3MM). The membrane was floated in a dish of ddH<sub>2</sub>O until it wetted completely from below. It was then immersed in alkaline transfer buffer for a minimum of 5 minutes. A triangle was cut from the bottom left hand corner of the membrane to allow subsequent orientation.

### 2.7.7.3 Assembly of transfer apparatus and transfer of DNA

DNA was transferred to the membrane by capillary blotting using downward transfer. The assembly of the transfer apparatus is shown in Fig 2.1. Briefly a glass dish was filled with alkaline transfer buffer, a glass plate placed on top and then a stack assembled (all cut to same size as gel) comprising: a 5-7cm stack of absorbent paper towels, 8 pieces of dry filter paper, 2 pieces of wet filter paper (soaked briefly in transfer buffer), the Hybond-XL membrane, the denatured agarose gel, another 2 pieces of wet filter paper then a wick of filter paper the same width as the gel with both ends immersed in the buffer reservoir in the glass dish and finally a gel tray with 400g weight on top. The stack was assembled carefully to ensure no air bubbles between the layers and no bridging between the layers above and below the membrane/gel interface. The blot was then left overnight to allow DNA transfer between the gel and the membrane.



**Fig 2.1 Assembly of Southern blot downward transfer apparatus**

The following day the blot was disassembled and the position of the wells and bottom left hand corner marked on the membrane with pencil. The DNA was then fixed to the membrane by soaking in neutralisation buffer II for 15 minutes at room temperature.

#### 2.7.7.4 Preparation of $^{32}\text{P}$ labelled DNA probe

Complementary DNA probes of 300-450bp were synthesised by PCR and TOPO cloned into a pCR4TOPO vector (Invitrogen) as described in section 2.12.1.1. To release the probe from this vector, 5µg of vector DNA were digested with appropriate restriction enzymes, the products were run on a gel, the appropriate sized band excised and gel purified (Qiaquick gel extraction kit) with elution in 50µl elution buffer (EB). A 10µl aliquot was then used in a  $^{32}\text{P}$  labelling reaction taking standard radioactive precautions:

---

10µl gel purified unlabelled probe added to 35µl TE in a screw-top microcentrifuge tube

Denature by boiling for 5 minutes then place on ice for 5 minutes

Spin down and add to Ready-to-go DNA labelling beads –dCTP (Amersham Biosciences)

Add 5µl (50µCi) [ $\alpha$ - $^{32}\text{P}$ ] dCTP (3000Ci/mmol) (Amersham Biosciences), mix by gently pipetting up and down

Incubate 37°C for 30 minutes.

Remove un-incorporated [ $\alpha$ - $^{32}\text{P}$ ] dCTP using a ProbeQuant G50 microcolumn (Amersham Biosciences) according to manufacturer's instructions

Denature labelled probe by boiling for 5 minutes, place on ice 5 minutes

Add 1µl of probe to 1 vial of scintillation fluid, and count on a scintillation counter (Beckman LS6500 Multipurpose Scintillation Counter).

---

#### 2.7.7.5 Probing of membrane for complementary DNA sequence

The fixed Hybond-XL membrane was wetted with 2x SSC and then placed DNA side up on to nylon mesh and into a hybridisation bottle. 0.1ml/cm<sup>2</sup> of prewarmed (50°C) hybridisation buffer - ULTRAhyb (Ambion) was added and the bottle was placed in a hybridisation oven (Hybaid mini-10 oven, Thermo Hybaid, Ashford, UK) with rotation for between 40 minutes and 4 hours. At the end of this pre-hybridisation period the buffer was replaced with the same volume of fresh ULTRAhyb and labelled probe was added to achieve 1x10<sup>6</sup> counts/minute per ml of buffer. The membrane was hybridised at 50°C overnight in a Hybaid oven with rotation.

The following morning, wash solutions were prewarmed to 55°C and then the radioactive hybridisation buffer was carefully discarded. The membrane was washed for 2x 5 minutes with low stringency wash buffer followed by 2x 15 minutes with high stringency wash buffer. The blot was then dried between two pieces of filter paper, wrapped in saran wrap and exposed to Kodak medical X-ray film (general purpose blue) in a film cassette. The

cassette was placed at -80°C overnight and after thawing the next day was developed using a KONICA medical film processor (SRX-101A). If images were indistinct then the blot was re-exposed to film for up to 7 days at -80°C.

#### **2.7.7.6 Stripping the Hybond-XL membrane**

In cases where repeat hybridisation with fresh probe was required the Hybond–XL membrane was stripped by immersion in a 0.1% SDS solution at boiling point. It was then shaken on a rotary platform for approximately 30 minutes until the solution had cooled. The membrane was then washed in 2xSSC prior to prehybridisation as described above.

### **2.8 Molecular Biology Methods: RNA**

In all methods involving RNA careful attention was paid to minimising RNA degradation by environmental RNases. All plasticware and glassware was either supplied guaranteed RNase free or was autoclaved using an RNase removal protocol. Diethyl pyrocarbonate (DEPC) treated water was used in all solutions and all surfaces, pipettes and non-autoclavable equipment were sprayed with RNAzap (Ambion) RNase decontamination solution. Filter tips were used throughout.

#### ***2.8.1 Isolation of Total RNA from cell lines***

##### **2.8.1.1 Using Trizol reagent**

RNA was extracted from cell lines using a commercial mono-phasic solution of phenol and guanidine isothiocyanate (Trizol reagent, Invitrogen). Disruption of cells was achieved by direct lysis on tissue culture plates for adherent cells (1ml Trizol added per 10cm<sup>2</sup> plate surface area). Suspension cells were pelleted by centrifugation (300g x 5 minutes) and 1ml of trizol added per 5x10<sup>6</sup> cells. Following addition of Trizol reagent, cells were homogenised by repeated pipetting. The homogenised samples were incubated for 5 minutes at room temperature to permit complete dissociation of nucleoprotein complexes. Phase separation was then performed by addition of 0.2ml of chloroform per 1ml Trizol reagent. Tubes were vigorously shaken then incubated for 2-3 minutes at room temperature. The samples were then centrifuged at 11000g x 15 minutes (at 4°C). Following this, the mixture separates into a lower red phenol-chloroform phase and an upper RNA containing aqueous phase. This colourless aqueous phase was transferred to a

fresh tube and RNA precipitated by the addition of 0.5ml isopropyl-alcohol per 1ml Trizol reagent. Following incubation for 10 minutes at room temperature the samples were centrifuged at 11000g x 10 minutes (at 4°C). After removal of the supernatant the RNA pellet was washed with 75% Ethanol (1ml per 1ml of Trizol), mixed by vortexing and centrifuged at 7500g for 5 minutes (at 4°C). The pellet was allowed to air-dry for 5-10 minutes before dissolving in an appropriate volume of RNase free water or TE buffer (usually 50µl). To evaporate any residual ethanol this solution was incubated at 60°C for 10 minutes.

### **2.8.1.2 Using RNeasy kit**

A second method using a RNeasy mini kit (Qiagen) was utilised to prepare RNA for subsequent use in quantitative PCR. This commercial kit uses a high salt buffer system in combination with ethanol to bind RNA to a silica membrane and uses spin-column technology to allow rapid washing and subsequent elution of high purity RNA. This procedure only isolates RNA greater than 200bp and therefore relatively enriches for mRNA as many ribosomal RNAs are smaller than this. Cells were counted and then pelleted by centrifugation at 300g x 5 minutes. The supernatant was fully removed and lysis buffer added (350µl for  $< 5 \times 10^6$  cells or 600µl lysis buffer for  $5 \times 10^6$ - $1 \times 10^7$  cells). The lysates were homogenised by passage 5-10 times through a blunt-ended 20 gauge needle. Samples could be stored at this stage at -80°C if required. If so then thawing was performed in a 37°C waterbath to ensure all salts were dissolved. Subsequent ethanol addition, binding to the membrane, and wash steps were performed according to the manufacturer's instructions. Following the initial wash step an on-column DNase digestion was performed to eliminate any genomic DNA contamination using an RNase-free DNase set (Qiagen) again according to manufacturer's instructions. After the final wash step an additional centrifugation step was performed to remove any residual traces of wash buffer before elution of the RNA in 30µl of RNase-free water.

### **2.8.2 Removal of contaminating DNA from RNA**

On column DNase digestion was performed for all RNA samples isolated with the RNeasy kit as described above. For other samples DNase digestion was performed using a commercial DNase enzyme: DNA-free (Ambion) as described:



---

**DNase treatment (perform in 0.5ml tube)**


---

Sample (up to 10µg RNA)	xµl
DNase I buffer (10x)	5 µl
Recombinant DNase I	2µl (1µl at outset, 1µl after 30minutes)
Water	To final volume of 50µl

---

Mix gently, incubate room temp 1 hour, add 5µl DNase inactivation reagent (resuspend thoroughly before use), mix well, incubate room temp for 2-3 minutes mixing several times, centrifuge at 10,000g for 1.5 minutes and transfer the RNA containing supernatant to a fresh tube

---

### **2.8.3 Quantification of RNA**

RNA concentration was estimated by measurement of  $A_{260}$ . Purity was assessed by calculation of the  $A_{260}:A_{280}$  ratio which should be  $\geq 1.8$  (when diluted in TE).

### **2.8.4 Reverse transcription of RNA**

Two commercial methods were used for generation of a cDNA from an RNA template. In early experiments Superscript First-Strand Synthesis (Invitrogen) was performed using oligodT primers to enrich for cDNA copies of mRNA, according to manufacturer's instructions. Later, and for all quantitative PCR, cDNA was made with random hexamer primers using AffinityScript multiple temperature reverse transcriptase (Stratagene) in the following reaction:

---

**Affinity Script Reverse Transcription:** To a nuclease free centrifuge tube add (in order):

---

2µg total RNA

300ng (3µl) random primers

xµl RNase-free water to a total volume of 14.2µl

Mix and incubate at 65°C for 5 minutes, slowly cool (over 10 minutes) to room temperature (to allow primers to anneal to RNA). Then add:

2 µl 10x Affinity Script RT buffer

2µl 100mM DTT

0.8µl 100mM dNTP mix

1µl Affinity Script reverse transcriptase

---

Pre-incubate 25°C for 10 minutes then incubate at 42°C for 1 hour. Inactivate by heating to 70°C for 15 minutes then place on ice for subsequent use as template in PCR

---

Where required, a second reaction was performed with all the above components except the reverse transcriptase to provide a minus (-) RT control. This was used to rule out genomic DNA contamination during RT-PCR.

## **2.9 Molecular Biology Methods: Protein**

### ***2.9.1 Protein extraction from cell lines and tissues***

Between  $1 \times 10^5$  and  $1 \times 10^7$  (typically  $1 \times 10^6$ ) cells were used for western blots. Suspension cells were pelleted by gentle centrifugation (300g x 5 minutes), washed in ice cold PBS, pelleted once more. Adherent cells were washed on the plate then harvested by mechanical scraping before pelleting as above. Murine tissues from section 2.6.2 were homogenised by passage through a 26 gauge needle prior to lysis. All cells then underwent lysis using either Laemmli 2x sample buffer diluted with an equal volume of PBS or using CellLyticM (Sigma) mammalian cell lysis/extraction reagent according to manufacturer's instructions.

### ***2.9.2 Preparation of nuclear and cytoplasmic extracts***

All reagents and tubes were pre-cooled on ice. Cells were harvested as above, followed by washing with 1ml ice-cold PBS. The suspension was then centrifuged 14,000rpm for 10 seconds and the pellet resuspended in 500µl of cold Buffer A. The mixture was placed on ice for 10 minutes, flicking the tube every few minutes to resuspend the cells. Then it was vortexed for 10 seconds and spun 14,000rpm for 10 seconds. The supernatant was aspirated, placed in a fresh tube and stored at -80°C until use – this is the cytoplasmic fraction. The pellet was then resuspended in 25-50µl of cold buffer C, depending on the desired final concentration. This reaction was placed on ice for 20 minutes, flicking the tube once or twice, followed by centrifugation 14,000rpm for 3 minutes at 4°C. This supernatant (the nuclear fraction) was transferred to a fresh tube and stored at -80°C until use.

### ***2.9.3 Separation of proteins on polyacrylamide gels***

21µl of the protein extracts from above (except those already in Laemmli buffer) were mixed with 7µl of 4x NuPAGE LDS sample buffer (Invitrogen). All samples were then heated to 70°C for 10 minutes. Samples were loaded onto a pre-cast 4-12% NuPAGE Novex Bis-Tris gradient gel (Invitrogen) in a vertical electrophoresis tank (Xcell Surelock,

Invitrogen) filled with NuPAGE MOPS SDS running buffer (Invitrogen) with addition of 100µl 1M DTT to the 200ml of buffer in the central chamber. For size determination either (i) Novex Sharp (ii) Magic mark or (iii) Multimark protein standards (all Invitrogen) were run in lanes 1 and 10. Electrophoresis was performed for 1-2 hours at 100-150V until the loading dye front reached the bottom of the gel.

### **2.9.4 Western blotting**

Following electrophoresis proteins were transferred to NuPAGE nitrocellulose membranes (Invitrogen) using a tank transfer method (Xcell II blot module). Briefly, 6 blotting pads were soaked in 1x NuPAGE transfer buffer (Invitrogen), 2 pieces of filter paper and the nitrocellulose membrane were briefly wetted in the same buffer. The gel was rinsed under running water, the plastic casing cracked open with a spatula and the gel trimmed. A piece of filter paper was placed on top of the gel, this was then inverted and the membrane placed on the other side ensuring no air bubbles are trapped, The membrane was covered with a second piece of filter paper and then sandwiched between the 6 blotting pads (3 either side) in the blot module with the gel closest to the cathode core. After insertion into the vertical tank the blot module was topped up with transfer buffer and the outside of the module cooled by filling the outside chamber with distilled water. A current of 150mA (low voltage) was applied for 1-2 hours.

Following transfer the membrane was blocked in 5% milk/PBS Tween 0.1% (PBST) at 4°C overnight (or 1-2 hours at room temp if required) with gentle agitation. Membranes were then incubated with primary antibody diluted in 5% milk PBST at optimal concentration (listed in table 2.13) for 1-2 hours. They were then rinsed and washed in large volumes of PBST with gentle agitation and frequent changes of wash buffer ie 2 x 5 minutes then 1 x 15 minutes and 1 x 30 minutes. This was followed by a 45-60 minute incubation with an HRP conjugated secondary antibody diluted in 5% milk/PBST at optimal concentration (table 2.13). A second wash step was then performed as above. Binding of antibody was detected by an HRP mediated chemiluminescent reaction using Supersignal West Pico chemiluminescent substrate (Pierce) according to manufacturers instructions. The membrane was then placed between acetate sheets and exposed to X-ray film. Films were initially exposed for 2 minutes and then the exposure times were varied between 15 seconds and 2 hours according to the intensity of any bands resulting from the initial exposure.

### 2.9.5 Estimation of protein loading

Where possible, equal starting amounts of cells were used for protein extraction. Equal protein loading was then confirmed either protein staining of the membrane with Ponceau S reagent (Sigma) or by stripping the membrane and reprobing with a housekeeping gene (actin or GAPDH). For Ponceau staining the membrane was covered with stain, rocked for 5 minutes at room temperature and then washed with distilled water until all background staining was removed. For stripping, membranes were incubated in western stripping buffer at room temperature for 1 hour before washing in PBST and proceeding to re-blocking followed by primary and secondary antibody incubations as described above.

## 2.10 Molecular Biology Methods: Polymerase Chain Reaction

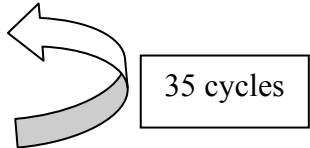
### 2.10.1 Standard PCR

PCR was performed using a variety of commercially available DNA polymerases depending on the downstream application. Oligonucleotide primers were designed either by eye or with the help of Primer3 express software (<http://frodo.wi.mit.edu>). All potential primer sequences were checked for secondary structure and predicted melting temperature using oligo.exe software. Primers were designed according to the following criteria:

Length	19-24 basepairs
GC content	Approximately 50%
Terminal G or C bases	1 or 2
Melting temp ( $T_m$ ) (to match as closely as possible for forward and reverse primer)	62-65°C
Location	Near 3' end of gene (for RT PCR) Span introns if possible (for RT PCR)
Secondary structure	No loop or loop with $T_m$ well below annealing temperature of primer pair

Primer sequences are listed in section 2.4. All primers were supplied lyophilised. Initially 100 $\mu$ M stocks were made by addition of ddH<sub>2</sub>O. A 1 in 10 dilution produced a 10 $\mu$ M working solution for use in PCR. Standard PCR was performed using commercially available pre-mixed master mix (comprising 1.1x buffer, MgCl<sub>2</sub>, 1mM dNTPs and Taq

DNA polymerase) – either ReddyMix (Abgene) or Red PCR Mastermix (Rovalab). These were supplied as 45µl pre-aliquoted PCR tubes to which 2.5µl of template cDNA/DNA and 1.25µl of each primer (forward and reverse 10µM solutions) was added. Certain downstream applications (WAVE, generation of targeting construct) required custom PCR reagents and set up and are discussed in the relevant sections (2.11.2 and 2.10.4 respectively). PCR was performed on a PTC-200 thermal cycler (MJ Research) using the following cycling parameters:

Segment	RT - PCR	PCR from genomic DNA
Segment 1 – initial denaturation	95°C for 2 minutes	95°C for 5 minutes
Segment 2 – thermal cycling	95°C for 30 seconds 55°C* for 30 seconds 68°C for 1 minute per kb	
Segment 3- final elongation		68°C for 10 minutes 4°C until use

\*Annealing temperature was varied between 52 and 58°C depending on the calculated  $T_m$  of the primer pair – initially a temperature was chosen that was 2-5°C below the predicted  $T_m$ . If this failed to produce a satisfactory single product then the temperature was titrated in 2°C increments (down for poor yield, up for non-specific bands) optimum annealing temperatures are listed alongside primer sequences in section 2.4.

### 2.10.2 *Semiquantitative PCR*

Semiquantitative PCR was performed by reducing the number of amplification cycles to 25 to ensure product accumulation is still in the exponential phase. Genes of interest were amplified alongside actin or GAPDH housekeeping gene controls. 20µl of product was run on an agarose gel with ethidium bromide. Gels were then photographed using a UV gel imager system. The amount of amplified product was estimated using Alpha-imager software to perform densitometric analysis of photographed bands. Results were normalised to densitometric readings of the corresponding housekeeping gene control.

### 2.10.3 *Nested PCR*

PCR screening for gene targeting events was performed using nested PCR. 30 cycles of PCR amplification were performed with external primers before transferring 5µl of this

reaction mixture to a fresh PCR tube containing mastermix with internal primers. This second PCR reaction was run for a further 35 cycles and the products analysed on a 1.2% agarose gel with ethidium bromide alongside size markers and a positive control.

#### **2.10.4 Proof-reading PCR**

Platinum *Pfx* proof reading DNA polymerase (Invitrogen) was used for generation of the ES cell targeting construct. Reaction parameters were optimised as described in chapter 4.4.2. Following optimisation, the following reaction conditions were used:

10x <i>Pfx</i> Amplification buffer	10µl (final concentration 2x)
10x PCRx enhancer solution	5µl
10mM dNTP	1.5µl
50mM MgCl <sub>2</sub>	1µl
Primer mix (10µM each)	1.5µl
Template	1µl
Platinum <i>Pfx</i> polymerase	1µl
Nuclease free water	29µl
Reaction components were mixed, spun down and cycled using standard genomic DNA cycling conditions (section 2.10.1)	

#### **2.10.5 Quantitative PCR using fluorescent probes**

Quantitation of relative transcript levels of the GATA-1 isoforms was performed using Taqman technology (Applied Biosystems). Primer and probe design were carried out by collaborators (Dr Oliver Tunstall-Pedoe, Imperial College, London) and sequences are listed in section 2.4.7. The assay was performed on an Applied Biosystems 7900HT Fast Real-Time system. All assays were performed in triplicate and normalised to ribosomal 18S RNA transcript levels also assayed in triplicate. A no template control was also run for each primer/probe pair. Assays were prepared as follows:

For each reaction prepare diluted template and a master mix containing primers and probes, all reactions were performed in triplicate	
Mastermix:	1 reaction
2x Universal PCR Master Mix (Applied Biosystems)	10µl
Forward primer (20µM)	0.3µl

---

Reverse primer (20 $\mu$ M)	0.3 $\mu$ l
FAM/TAMRA probe (10 $\mu$ M)	<u>0.4<math>\mu</math>l</u>
Total volume	11 $\mu$ l
Template mix	1 reaction
cDNA (diluted 1 in 5 following RT reaction using 2 $\mu$ g RNA)	2 $\mu$ l
Nuclease free water	<u>7<math>\mu</math>l</u>
Total	9 $\mu$ l

---

To perform an assay in triplicate 3x template mix (27 $\mu$ l) and 3x master mix (33 $\mu$ l) were transferred to a fresh microcentrifuge tube, vortexed gently to mix, spun down and 20 $\mu$ l was transferred to each of 3 wells on a 96 well thin-walled PCR plate. Once all the samples were loaded in this way, the plate was sealed, spun down briefly and run on 7900HT thermal cycler using standard reaction conditions: initial denaturation 94°C x 10 minutes, then 40 cycles of 15 seconds at 95°C followed by 60 seconds annealing/extension at 60°C, with fluorescence measured at the end of each extension step.

---

Control 18SrRNA reagents were supplied as a 20x primer/VIC/TAMRA probe mix (Applied Biosystems). For a single reaction, 1 $\mu$ l was added to 10 $\mu$ l 2x universal PCR Master Mix and then mixed with 9 $\mu$ l template mix, as described above, to produce a final reaction volume of 20 $\mu$ l.

Results were analysed using Sequence detection systems (SDS) software v2.1.

### **2.10.6      *Use of Taqman low-density array plates***

To simultaneously quantitatively assay 15 genes of interest alongside an 18SrRNA control a custom Taqman low density array (Applied Biosystems) was designed using ABI software ([www.appliedbiosystems.com](http://www.appliedbiosystems.com)). This consists of a 384 well micro-fluidic card preloaded with validated primer and probe pairs (in triplicate) for each gene. cDNA was mixed with PCR master mix and a loading port used to add the reaction mixture evenly to all the wells for that particular sample. 8 biological samples were analysed per 384 well card (8 samples x 16 genes x 3 = 384). To reduce variability due to sample processing all RNA was extracted on the same day, quantification of RNA by spectrophotometry was performed in a single batch and 2 $\mu$ g of RNA were used in cDNA synthesis again in a single batch. The assay was set up as follows:

---

Dilute cDNA (made using random primers) 1 in 10 (acceptable range 30-1000ng)
Add 50 $\mu$ l diluted cDNA to 50 $\mu$ l Universal PCR Master Mix (Applied Biosystems)
Gently vortex and spin down

---

---

Load 100µl into the fill port, spin 1200rpm x 1 minute (twice) to fill the wells

Seal the plate and cut off the loading ports with scissors

Run on 7900HT PCR system using low density array thermal cycling block using standard cycling conditions: Initial denaturation 94.5°C x 10 minutes then 40 cycles of 97°C x 30 seconds followed by 59.7°C x 1 minute.

Analyse results using SDS v2.1 software

---

## 2.11 Molecular Biology Methods: Mutation detection using DHPLC (WAVE)

### 2.11.1 *PCR set-up for exon amplification*

WAVE primers and PCR conditions had been previously optimised by our collaborators (P.Vyas, C.Fisher WIMM, Oxford). Primer pairs were split in to two groups – those requiring buffer A and those requiring buffer B as detailed in section 2.4.5. These buffers were identical except for the addition of 7µl 5M Betaine to buffer B (to reduce secondary structure formation) as shown below:

	Buffer A	Buffer B
10x Applied Biosystems PCR buffer II	5µl	5µl
25mM MgCl <sub>2</sub>	3µl	3µl
25mM dNTP mix	0.4µl	0.4µl
Amplitaq Gold DNA polymerase (Applied Biosystems)	0.2µl	0.2µl
Pwo polymerase (Roche)	0.05µl	0.05µl
Nuclease free water	38.35µl	31.35µl
5M Betaine	n/a	7µl

To buffer mix add 1µl of forward and 1µl reverse primers (10µM) plus 1µl template (100ng genomic DNA)

Cycling conditions: (i) initial denaturation 95°C for 10 minutes (ii) 35 cycles of 95°C 1 minute, annealing temp (see table 2.9) 1 minute, extension 72°C 1 minute (iii) final extension 72°C 5 minutes (iv) store at 4°C until ready to use.

---

Check amplicon by electrophoresis on 2% agarose gel 100V for 1 hour

---



### **2.11.2      *WAVE Heteroduplex conditions***

10µl test PCR reaction and 10µl of HEK293 control reaction were mixed in a WAVE 96-well plate. Plates were sealed with an adhesive lid and centrifuged briefly, then run on an ABI thermal cycler, using the following conditions:

---

Denature 95°C for 5 minutes

Ramp down to 68°C at 0.1°C per second

Maintain at 68°C for 5 minutes

Ramp down to 64°C at 0.1°C per second

Maintain at 64°C for 1 minute

Ramp down to 60°C at 0.1°C per second

Maintain at 60°C for 1 minute

Ramp down to 55°C at 0.1°C per second

Maintain at 55°C for 1 minute

Ramp down to 20°C at 0.1°C per second

Maintain at 20°C for 5 minutes

---

Centrifuge briefly and place on WAVE platform, inject 5µl for separation of homo and heteroduplexes at two different temperatures (as predicted by melt profile software)

---

## **2.12 Molecular Biology Methods: Vectors**

### **2.12.1      *Insertion and manipulation of vector DNA***

#### **2.12.1.1      TOPO cloning**

TOPO cloning exploits the fact that Taq DNA polymerases possess non-template dependent terminal transferase activity that adds single deoxyadenosine overhangs to the 3' end of PCR products. Therefore, if the foreign DNA segment has been produced using Taq polymerase it can be directly introduced into a linearised plasmid vector which contains complementary unpaired 3' deoxythymidine at the two ends. The TOPO cloning reaction is performed as follows:

---

Fresh PCR product 4µl

Salt solution 1µl

TOPO vector 1µl

---

Mix gently and incubate for 5-30 minutes at room temperature (30 minutes for PCR products >1kb)

---

The reaction is then placed on ice and used to transform One-Shot cells as described in section 2.12.2 below

### ***Addition of 3'A overhangs to PCR products***

In cases where PCR had been performed with a non-Taq polymerase (e.g. Pfu polymerase) 3' deoxyadenosine can be added to the final product using the following method:

---

Gel purify or PCR purify product (QiaQuick kit) and elute in 30µl EB. Take 7µl of this and add:

1µl 10x Taq polymerase buffer

1µl dATP

1µl Taq DNA polymerase (New England Biolabs)

---

Incubate 10-15 minutes at 72°C, use 4µl in TOPO cloning reaction as described above.

---

## **2.12.1.2 DNA Ligation**

Ligation reactions were used to join plasmid and insert DNA that were either both blunt-ended or that had been digested with restriction enzymes to produce compatible sticky ends. 5µg of vector and insert DNAs were digested in a 35µl reaction volume with 20 units of enzyme, 3.5µl of 10x RE buffer and ddH<sub>2</sub>O to 35µl. The products were run on a gel and the desired band excised and then purified (QIAquick kit, section 2.7.5). Where the vector contained internally compatible sticky ends, the vector fragment was dephosphorylated to prevent vector religation in the following reaction:

---

Plasmid DNA 30µl

Shrimp Alkaline Phosphatase (SAP) 1µl (Roche)

10 x SAP buffer 4µl (Roche)

ddH<sub>2</sub>O 5µl

---

Incubate at 37°C for 30 minutes, heat inactivate 70°C for 10 minutes and purify with QIAquick PCR purification kit

---

3 ligation reactions were set up for each pair: (i) vector plus insert (ii) vector alone (iii) insert alone. A commercial rapid DNA ligation kit (Roche) was used according to the following method:

Reagent	Vector + Insert	Vector only control	Insert only control
5x DNA dilution buffer	2µl	2µl	2µl
Vector DNA	2µl	8µl	0µl
Insert DNA	6µl	0µl	8µl
Add 10µl 2x ligase reaction buffer and 1µl DNA ligase to each tube, mix gently and incubate at room temperature 10-15 minutes			
Use 2µl of ligation reaction to transform DH5α cells (section 2.12.2)			

In difficult reactions the proportions of vector to insert were varied between 1:1 and 1:10. Following growth overnight on selective plates, the numbers of colonies on test and control plates were compared. The ratio of test:vector-only colonies allowed an estimation of the rate of vector religation. If religation rates were high 24 colonies were picked, if low then approximately 10 colonies were picked and processed as described in section 2.12.2.3.

### ***Blunt ended ligation***

In cases where no compatible sticky ends could be identified, vector and insert were either cut with restriction enzymes that produce blunt ends, or protruding 5' or 3' termini were blunted by the use of T4 DNA polymerase. This enzyme removes 3' overhangs via its 3'-5' exonuclease activity and fills in 5' overhangs to form blunt ends. The following reaction was performed:

5x T4 reaction buffer 4µl
Digested DNA 1µg
dNTP mix (2mM each) 1µl
T4 DNA polymerase 0.2µl (1 unit)
Nuclease free water to 20 µl
Incubate for 5 minutes at room temperature, then stop the reaction by heating to 75°C for 10 minutes

### 2.12.1.3 Site-directed mutagenesis

To introduce targeted mutations into plasmids carrying a DNA sequence of interest commercial mutagenesis kits were used (QuikChange II and QuikChange XL site directed mutagenesis kits, Stratagene). Two oligonucleotide primers were designed containing the desired mutation according to the following criteria: (i) annealing to the same sequence on opposite strands (ii) 25-45 basepairs in length (iii) mutation should be in the middle of the primer with 10-15 basepairs of normal sequence either side (iv) minimum GC content of 40% and end in  $\geq 1$  G or C bases (v) purified by HPLC. Primer sequences are listed in table 2.6. Primers annealed to plasmid DNA and were extended during a thermal cycling reaction using a high fidelity DNA polymerase as detailed below. At the end of the procedure native (un-mutated) plasmid DNA was degraded by an endonuclease (Dpn1) which only recognises methylated DNA (almost all *E.coli* strains used to produce plasmid preparations *Dam* methylate DNA and therefore mark it for degradation by this enzyme). Following digestion of the parent DNA, the mutated vector DNA was transformed into XL1-Blue supercompetent cells, spread on selective plates and colonies picked the next day. Screening of colonies for those that have incorporated the mutation was performed by taking advantage of changes in restriction enzyme sites introduced by the mutation (as described in results chapter 4). Mutations were confirmed by plasmid DNA sequencing. For plasmids <8kB the QuikChange II kit was used, for greater than 8kB QuikChange XL was used. The two methods are similar apart from the addition of Quikchange solution and the use of a highly competent bacterial strain (XL1-Gold) to optimise transformation efficiency for larger plasmids. The mutagenesis reaction is detailed below:

---

In a thin-walled 0.2ml PCR tube add the following:

---

5 $\mu$ l 10x reaction buffer

10ng (x $\mu$ l) dsDNA template

x $\mu$ l (125ng) of primer 1, x $\mu$ l (125ng) of primer 2

1 $\mu$ l dNTP mix (proprietary)

(for QuikChange XL add 3 $\mu$ l QuikSolution)

ddH<sub>2</sub>O to a final volume of 50 $\mu$ l

Then add 1 $\mu$ l PfuUltra HF DNA polymerase (2.5U/ $\mu$ l)

Denature 95°C for 30 seconds then perform 12 cycles comprising: 95°C 30 seconds, 55°C 1 minute and 68°C for 1 minute per Kb of plasmid

---

Add 1 $\mu$ l of Dpn1 restriction enzyme directly to the amplification products, mix by pipetting up and down, spin and incubate at 37°C for 1 hour.

---

Use 1µl of this Dpn1 digested product to transform XLI-Blue (QuikChange II) or XL1-Gold (QuikChange XL) competent cells (as described in section 2.12.2).

## 2.12.2 *Bacterial Transformation with vector DNA*

### 2.12.2.1 Transformation of bacterial strains

Chemically competent cells were stored at -80°C and single-use aliquots (50µl) thawed on ice prior to use. Cells were then transferred to pre-chilled Falcon 2059 tubes and 2µl of vector DNA were added and gently mixed. The mixture was then incubated on ice for 5-30 minutes before being heat-shocked in a 42°C water bath for 20-45 seconds (without shaking) depending on the bacterial strain as detailed below:

Strain	Incubation time on ice	Heat shock	Additional notes
One shot TOP 10 (Invitrogen)	5-30 minutes	30 seconds	
XL1-Blue (Stratagene)	30 minutes	45 seconds	Add 1.7µl β-mercaptoethanol to thawed aliquot and incubate on ice 10 minutes before adding DNA
XL1-Gold Ultracompetent cells (Stratagene)	30 minutes	30 seconds	Add 2µl β-mercaptoethanol to thawed aliquot and incubate as above
DH5α library efficiency (Invitrogen)	30 minutes	45 seconds	
DH5α subcloning efficiency (Invitrogen)	30 minutes	20 seconds	

After heat-shocking, cells were placed on ice for 2 minutes and then 250µl of prewarmed SOC media (without antibiotic) were added. Tubes were capped tightly and shaken horizontally (200rpm) at 37°C for 1 hour. 10-50µl were then spread onto an LB Agar plate pretreated with an appropriate selective antibiotic (usually Ampicillin or Kanamycin) and/or IPTG and X-gal if blue white colour screening is required. (see 2.12.2.2 below). Plates were inverted and incubated at 37°C overnight.

### 2.12.2.2 Selection strategies

A number of different strategies were employed to select for bacteria that were transformed with the DNA sequence of interest. Firstly bacterial cells that had taken up plasmid DNA were selected by growth of bacteria in nutrient medium containing an antibiotic normally effective against that strain. The plasmid encodes an antibiotic resistance gene and therefore only bacterial cells that contain plasmid DNA are able to survive and proliferate. To further select for plasmids that had incorporated foreign DNA two methods were employed:

#### ***Blue-white colour screening***

Some vectors (e.g. pCR2.1TOPO and pBluescript II) contain a  $\beta$ -galactosidase gene whose promoter and coding sequence are interrupted by the multiple cloning site. Vectors that contain insert will be unable to produce  $\beta$ -galactosidase (or only at very low efficiency) whereas those that have religated without an insert will produce active enzyme. Plates are prepared by spreading the  $\beta$ -galactosidase substrate X-gal (5-Bromo-4-chloro-3-indolyl  $\beta$ -D-galactopyranoside) (40 $\mu$ l of 40mg/ml stock solution) and IPTG (Isopropyl-thio- $\beta$ -D-galactosidase) (40 $\mu$ l of 100mM stock) on to the plate prior to plating of bacteria. Colonies containing  $\beta$ -galactosidase activity are stained dark blue. For these vectors, only white or pale blue colonies were picked for further analysis.

#### ***Use of lethal genes***

The pCR4TOPO vector has a lethal E.coli gene *ccdB* fused to the C-terminus of the Lac-Z sequence. This gene is disrupted by the presence of insert but is lethal in cases where the plasmid has religated without insert. In this way only colonies containing recombinants should survive plating on antibiotic selective agar.

### 2.12.2.3 Isolation and purification of vectors from bacteria

Following growth of transformed bacteria on selective plates, individual colonies were picked with sterile pipette tips and grown up in 3mls (minipreps) or 100mls (maxipreps) of nutrient medium containing selective antibiotic prior to isolation of plasmid DNA:

#### ***Small-scale isolation of plasmid DNA (miniprep)***

Minipreps were generally performed using a commercially available kit (QIAprep Miniprep, Qiagen). This method is based on alkaline lysis of bacterial cells followed by adsorption of DNA to a silica membrane in high salt conditions. This is followed by

washing and subsequent elution of plasmid DNA. No phenol, chloroform or alcohol precipitation is involved.

Minipreps were performed on bacterial cultures grown from single colonies picked from selective plates. Colonies were cultured in 3mls of LB medium containing the appropriate antibiotic, shaken horizontally at 200rpm 37°C overnight. The cells were then harvested by centrifugation at 7500g for 3 minutes using a standard benchtop microcentrifuge. The cells were resuspended in 250µl of lysis buffer (buffer P1) and extraction of DNA performed according to manufacturer's instructions. DNA was eluted in 30µl EB buffer.

Isolation of DNA from large plasmids (>10kb) was also performed using commercial reagents (Qiagen Mimiprep kit) but spin columns were not used as these significantly reduced DNA yield. Instead cell pellets were resuspended in 150µl buffer P1 followed by addition of 150µl of buffer P2. Following mixing by inversion, 250µl of buffer P3 was added. The samples were spun at full-speed on a bench-top centrifuge for 10 minutes. The supernatant was then mixed with 1ml of 100% ethanol and spun for 15 minutes at full speed. The ethanol was decanted and the pellet resuspended in 30µl of EB buffer.

In some cases, where screening of multiple colonies with large inserts was required, colonies were picked using sterile pipette tips and the tip ejected into a well of a 96-well tissue culture plate containing 100µl LB medium with selective antibiotic. The plate was sealed and shaken at 200rpm overnight. The following day 20µl of the overnight culture were transferred to a fresh 96-well plate (test plate) and mixed with 50µl of pre-warmed plasmid lysis buffer (with bromophenol blue). The original plate (master plate) was placed at 4°C to allow later expansion of any positive colonies. The test plate was incubated at 37°C for 5 minutes then placed on ice for 5 minutes. It was then spun at 1200rpm for 5 minutes and 20µl the loaded directly onto an agarose gel, run alongside the native vector and a size marker. Colonies containing the desired insert were identified by a size change. They were then grown up from the master plate, mini or maxi-prepped and their identity checked by restriction enzyme digestion or sequencing.

### ***Large scale isolation of plasmid DNA (maxiprep)***

To produce larger amounts of endotoxin-free plasmid suitable for transfection into mammalian cell lines, DNA was prepared using the Qiagen endo-free plasmid Maxi kit (Qiagen). This kit employs a different principle than the mini-prep kit with a modified alkaline lysis protocol followed by endotoxin removal and then binding of plasmid DNA to an anion-exchange resin under low-salt conditions. The resin is then washed in medium

salt conditions, and plasmid DNA eluted in a high salt buffer before isopropanol precipitation and ethanol washing of the DNA pellet.

Single colonies were picked from selective agar plates, grown in 5ml LB at 37°C with shaking for 8 hours then transferred to 100ml of LB media – followed by overnight incubation at 37°C with horizontal shaking. Bacteria were then harvested by centrifugation at 6000g for 15 min at 4°C (using a Sorvall RC5B centrifuge with a GS3 rotor). The pellet was resuspended in 10ml of lysis buffer (Buffer P1) and DNA extraction performed according to manufacturers instructions (for plasmids <8kb). Following isopropanol precipitation and ethanol washing of the DNA pellet (as described in manufacturer's protocol) the pellet was air-dried for 5-10 minutes and resuspended in 200µl of endotoxin-free TE buffer.

For plasmids >8kB the Qiagen kit was still used but samples were filtered through sterile gauze rather than using the Qiafilter cartridge, prior to passage through an equilibrated QIAGEN-tip 500 anion-exchange resin. In addition, elution buffer was warmed to 50°C prior to use. All other steps were as directed in the manufacturer's protocol.

### **2.12.3      *Transfection of vectors into mammalian cell lines***

The majority of routine transfections in adherent cells were performed using the commercial lipid-based transfection reagent Fugene 6 (Roche). In some cases alternative reagents were used such as Effectene (Qiagen), Polyfect (Qiagen) and Lipofectamine (Invitrogen). Suspension cells were more commonly transfected by nucleoporation using an Amaxa system. All these methods are described below. SiRNA was transfected into suspension cultures using Lipofectamine 2000 (Invitrogen) as discussed in section 2.13.7.

#### **2.12.3.1      General considerations**

Cells were routinely passaged 24 hours prior to transfection and fed on the day of transfection to ensure they were in log-growth phase. Plasmid DNA was prepared using a commercial maxiprep kit (Qiagen see section 2.12.2.3) with an additional endotoxin removal step (maxiprep – endo-free). For stable transfections, plasmid DNA was linearised with a suitable restriction enzyme, gel-purified and ethanol precipitated (with resuspension in sterile TE or water in a tissue culture hood) prior to transfection. Transient transfections were assayed 48-72 hours post transfection. Stable transfections were placed in selective media at 24 hours post transfection and kept in selective media even after establishment of



stable clones. Depending on the downstream application either individual clones were picked and expanded up, or pools of stable transfectants were used, as described in the relevant chapters.

For stable transfections in previously untested cell lines antibiotic titration was performed to determine the minimum concentration of antibiotic required to kill all untransfected cells. Serial dilutions of antibiotic (within its published working range) were added to cell culture medium in 6-well plates and untransfected cells seeded at a density of  $1 \times 10^5$  cells/ml. Cultures were observed daily for cell death. The lowest concentration of antibiotic to cause 100% cell death after 1-2 weeks in culture was used in subsequent selection experiments.

### **2.12.3.2 Transfection with Fugene 6 reagent**

Optimisation was performed for each individual cell line using Fugene 6:DNA ratios of 3:1, 4:1, 3:2 and 6:1 in 6 well plates. The majority of assays were performed using 10cm tissue culture plates. All quantities are given for transfection of a single 10cm plate. The reaction was scaled up or down depending on the individual experiments. Scaling ratios are given in the manufacturer's instructions for use. For a standard transfection using a 4:1 ratio of Fugene 6:DNA:

---

Add 20 $\mu$ l of Fugene 6 directly into 550 $\mu$ l of serum-free medium - do not allow the undiluted Fugene 6 to come into contact with plastic surfaces. Vortex for 1 second and incubate for 5 minutes at room temperature

Add 5 $\mu$ g DNA to the above (max volume 50 $\mu$ l), vortex 1 second, incubate 15-30 minutes at room temperature

Add the mixture drop by drop to the cells to be transfected, swirling the dish to ensure even distribution.

---

Incubate under standard cell culture conditions, assay for gene expression at 48-72 hours (transient) or place in selective antibiotic at 24 hours for stable transfectants.

---

### **2.12.3.3 Transfection with other reagents**

In certain cases alternative transfection reagents (as described in the relevant results sections) were used. The methods of transfection were similar to Fugene 6. These reagents were used according to the handbooks supplied with the kit with no deviations from protocol.

### 2.12.3.4 Nucleoporation

K562 and Meg-01 cells were transfected using a commercial nucleoporation system developed by Amaxa. Nucleofector kit C (Meg01) and kit V (K562) were used according to manufacturer's instructions. All nucleoporation was performed on an Amaxa Nucleofector device. Briefly:

---

Passage cells 2 days prior to transfection and aim for a density of  $2 \times 10^5$  cell per ml

Prepare 2 $\mu$ g of DNA in 1-5 $\mu$ l of sterile TE for each sample, add 1.5ml warmed culture media to each well of a 12-well plate

Add 0.5ml supplement to 2.25ml Nucleofector solution prior to use, then pre-warm.

Count cells and centrifuge  $1 \times 10^6$  per nucleofection sample at 1200rpm for 5 minutes

Discard the supernatant and resuspend in warmed Nucleofector solution at a density of  $1 \times 10^6$  cells/ml

Perform each nucleofection separately: mix 100 $\mu$ l of cells with 2 $\mu$ g DNA, transfer to an Amaxa cuvette and close cap. Insert into machine and lock in position

Select programme T-16 (K562) or X-05 (Meg01) and press X key to start

---

Remove samples from the cuvette immediately by adding 500 $\mu$ l of pre-warmed media and transferring to the pre-prepared 12-well plate using a plastic pipette. Incubate at 37°C.

---

### 2.12.3.5 Electroporation

Transfection of large plasmid vectors into murine ES cells was achieved by electroporation as follows:

---

Wash cells in PBS and resuspend in sterile PBS at a concentration of  $5 \times 10^6$  cells per 700 $\mu$ l

Add 5-10 $\mu$ g linearised plasmid DNA and mix gently, avoid air bubbles

Transfer to a 4mm electroporation cuvette, and place on ice

Dry cuvette and place in Genepulser II device (Bio-rad), electroporate at 250V/cm, capacitance 500 $\mu$ F. For transient transfections leave the cuvette in the electroporation chamber for 1-2 minutes before removing

---

Transfer cells to ice then to a 10cm plate and add 10ml pre-warmed culture medium

---

## **2.12.4 Assessment of efficiency of transfection**

### **2.12.4.1 $\beta$ -galactosidase assay**

A control vector carrying the bacterial *lacZ* gene (encoding  $\beta$ -galactosidase) under the control of the SV40 early promoter and enhancer (pSV $\beta$ gal) was used to monitor transfection efficiency. For adherent cells, in situ  $\beta$ -galactosidase staining was performed to assess the proportion of cells that had taken up plasmid DNA. Transfection efficiencies in suspension cells and quantitative estimation of transfection efficiency were obtained by using a  $\beta$ -galactosidase assay kit measuring enzyme activity in whole cell lysates.

#### ***$\beta$ -Gal staining kit***

To detect the proportion of cells that had taken up plasmid DNA, cells were fixed and then stained in the presence of X-gal using a commercially available kit ( $\beta$ -gal staining kit (Invitrogen)). X-gal hydrolysis is catalysed by  $\beta$ -galactosidase and produces a visible blue colour change. All quantities described are for staining a single 60mm plate. Cells were initially washed in 2.5ml PBS, followed by fixation with 3ml 1x fixative solution for 10 minutes at room temperature. Meanwhile the staining solution was prepared by adding 25 $\mu$ l each of solutions A,B and C to 125 $\mu$ l of X-gal (20mg/ml in dimethylformamide) and 2.3ml 1x PBS. After the 10 minute incubation the plate was rinsed twice with 2.5ml 1x PBS. 2.5ml of the staining solution was added to the plate and it was then incubated at 37°C for 30 minutes to 2 hours. Transfection efficiency was assessed by counting the number of blue cells versus the total number of cells in 5-10 random fields using 200x total magnification.

#### ***$\beta$ -Gal assay kit***

A quantitative estimation of  $\beta$ -galactosidase expression was obtained by measuring the yellow colour change produced by  $\beta$ -galactosidase mediated hydrolysis of the substrate ONPG (*ortho*-Nitrophenyl- $\beta$ -galactoside). Cells were washed twice with PBS and then lysed with reporter lysis buffer (RLB) (Promega). For adherent cells RLB was added directly to the tissue culture plate (400 $\mu$ l for a 60mm plate and 900 $\mu$ l for a 100mm plate) and the plate rocked a couple of times to ensure even distribution of reagent. The plate was then incubated at room temperature for 15 minutes at the end of which cells were removed using a cell scraper and transferred to a microcentrifuge tube placed on ice. They were then vortexed for 10-15 seconds and centrifuged at top speed for 2 minutes at 4°C. The supernatant was transferred to a fresh tube and stored at -80°C until use.

$\beta$ -galactosidase assays were performed in 96-well plates and a standard curve was run alongside the samples to ensure results were in the linear range, untransfected cells were run as a negative control to control for endogenous  $\beta$ -galactosidase activity and background absorbance. A commercial assay was used ( $\beta$ -galactosidase enzyme assay system #E2000, Promega), according to manufacturer's instructions and is summarised below:

---

Mix 30 $\mu$ l cell lysates with 20 $\mu$ l 1x RLB and transfer to a 96-well plate

Add 50 $\mu$ l 2x assay buffer to each well and mix well by pipetting

Cover and incubate at 37°C for 30 minutes – 3 hours until faint yellow colour appears

Stop the reaction by adding 150 $\mu$ l 1M sodium carbonate, mix well but avoid air bubbles

Read absorbance of the samples at 420nm (410-430nm) in a plate reader

---

For standard curve plot the absorbance at 420nm versus the concentration of  $\beta$ -galactosidase and ensure sample values are on the linear portion of the curve, if not repeat assay with more concentrated or dilute samples as needed

---

#### **2.12.4.2 Estimation of transfection efficiency using GFP**

Cells transfected with an internal ribosomal entry site (IRES) green fluorescent protein (GFP), IRES-GFP vector were examined directly for GFP expression using a fluorescent microscope (Axiostar plus with HBO50/AC mercury lamp, Zeiss) with image analysis using Axiovision MR software. An estimation of transfection efficiency could be obtained by counting the number of cells per high powered field. To give a more accurate estimation GFP negative cells could also be enumerated by counterstaining nuclei with DAPI and viewing under a confocal microscope (as described in section 2.13.6). Alternatively GFP expression levels could be quantitated by flow cytometry using the FL1 channel as described in section 2.13.2.

#### **2.12.5 Cell sorting for GFP expression**

Cell sorting for high GFP expression was performed with the help of Dr Ashley Gilmour – Glasgow Biomedical Research Centre.  $2 \times 10^6$  cells were washed in PBS, resuspended in FACS buffer with 1% fetal calf serum and passed through a cell strainer into a FACS tube before being placed on ice. Cells were sorted on a FACSaria (Becton Dickinson) instrument. Cells were initially gated on forward scatter (FSC) and side scatter (SSC) and then gated on high GFP expression. Cells in this sorting gate were diverted into the

collection stream and collected in FACS tubes containing 1.5ml of media. Where possible, 100,000 cells were collected. Following sorting, cells were counted and replated in an appropriate sized tissue culture vessel (usually a 24-well plate) to produce an optimum seeding density of  $2 \times 10^5$  cells per ml.

## 2.13 Cell biology methods

### 2.13.1 *Luciferase assay*

The pGL3hGATA-1 luciferase vector was a kind gift from Dr Paresh Vyas, WIMM, Oxford and is described in chapter 5 (5.2.4.3). NIH3T3 cells were co-transfected with a test plasmid (or positive/negative control plasmids), pGL3hGATA-1 and the  $\beta$ -galactosidase vector pSV $\beta$ -gal (# E1081, Promega). Cells were seeded at  $4 \times 10^5$  cell per 6cm tissue culture plate the day prior to transfection. Optimisation experiments (as described in chapter 5 section 5.2.4.3) led to the use of 1.5 $\mu$ g pGL3hGATA-1, 300ng pSV $\beta$ -gal and 1.5 $\mu$ g test plasmid, with a 4:1 ratio of Fugene 6 to DNA. Cells were lysed in situ with reporter lysis buffer (900 $\mu$ l per 10cm plate) and 30 $\mu$ l was used in a  $\beta$ -galactosidase assay (see 2.12.4 above) and 20 $\mu$ l in a commercial luciferase assay system (#E4030, Promega) read on a Fluoroskan Ascent FL Luminometer (Thermo Lab Systems). The assay was performed according to manufacturer's instructions, briefly:

---

Prepare luciferase assay reagent by the addition of 10ml luciferase assay buffer to the vial containing the lyophilised luciferase assay reagent, mix gently

Add 20 $\mu$ l cell lysates per well of a 96-well luminometer plate, place in the luminometer

Prime the injector port and programme the luminometer to add 100 $\mu$ l of luciferase assay reagent per well just prior to reading

Measure light produced for a period of 10 seconds

Move on to next well for a repeat of the injection-read process

---

Results were expressed as luminescence normalised to  $\beta$ -galactosidase activity in arbitrary units. A positive control (pGL3Luc control vector) and a negative control (untransfected cells) were run with each assay.

### 2.13.2 *Flow cytometry*

Cells were prepared for flow cytometry as follows:

---

Harvest  $1 \times 10^4$ - $1 \times 10^7$  cells (usually  $1 \times 10^6$ ), spin and wash in 5mls PBS

Resuspend cell pellet at a concentration of  $2 \times 10^5$ /100 $\mu$ l in BDFACS flow buffer (Becton Dickinson) with 1% FCS

If staining required: Take 100 $\mu$ l, place in FACS tube and add 5 $\mu$ l of human Fc receptor blocker, incubate 2-5 minutes at room temperature

Then add 5 $\mu$ l of each antibody gently flick to mix and incubate room temperature for 10 minutes

Add 2-3mls of FACSflow buffer/1%FCS to each tube, spin 1200rpm for 5 minutes

Pour off supernatant leaving about 200 $\mu$ l residual fluid

---

Analyse on FACS instrument

---

Flow cytometry was performed using FACSCalibur (Becton Dickinson), FACScan (Becton Dickinson) or MACSQuant (Milltenyl Biotec) instruments. Data were acquired from 10-20,000 gated events per sample and analysed using Cell Quest Pro software or MACSQuant software. In cases where more than one fluorochrome was being analysed compensation was performed with the help of experienced personnel.

### **2.13.2.1 Propidium iodide staining for DNA ploidy**

Ploidy analysis was performed using the following procedure:

---

Wash  $1 \times 10^6$  cells in cold PBS, spin and resuspend in 200 $\mu$ l cold PBS and vortex

In a 15ml centrifuge tube vortex 4mls ice-cold ethanol and add the cell suspension dropwise.

Incubate for a minimum of 45 minutes (overnight if possible)

Spin down, 1200rpm 10 minutes, aspirate supernatant and resuspend in 2ml PI mastermix, incubate 37°C for 30 minutes

Filter through a 40-70 $\mu$ m cell strainer

---

Analyse on FACSCalibur or MACSQuant instruments

---

A known 2N cell line (Raji) was used as a control and run alongside test samples

### **2.13.3 *In vitro* differentiation with TPA**

$1 \times 10^6$  cells were seeded into each well of a 6-well plate along with 3ml of tissue culture media. The phorbol ester TPA (12-O-Tetradecanoylphorbol-13-acetate) was added to a

final concentration of 100nM (100µM stock, diluted 1 in 1000 in culture media i.e. 1µl per ml culture media).

#### **2.13.4 *In vitro differentiation with Haemin***

Cells were prepared as above. A 4mM Haemin stock was prepared as described in section 2.1 and filter sterilised prior to use. Haemin was added to a final concentration of 10µM by addition of 2.5µl of 4mM stock per 1ml of culture media.

#### **2.13.5 *Cytospin and staining of suspension cells***

10<sup>5</sup> cells were washed twice in cold 2% FCS-PBS and then resuspended in 200µl of the same and placed on ice. A labelled glass slide, cardboard filter and sample chamber were then assembled, clipped together and placed in the cytospin rotor (Cytospin 3, Shandon). The sample was quickly aliquoted into the chamber and spun at 200rpm for 3 minutes. The filters were then removed, taking care not to contact the cell smears, and the slides allowed to air dry. Cells were then stained with May-Grunwald-Giemsa stain as described below:

---

Places slides in May-Grunwald stain for 5 minutes

Place slides in phosphate buffer pH 7.2 (Sigma-Aldrich) for 1.5 minutes

Place slides in dilute Giemsa solution (1:20 in deionised water) for 15-20 minutes

Rinse slides briefly in deionised water

---

Air dry, mount and evaluate

---

#### **2.13.6 *Confocal microscopy***

Confocal microscopy was performed with the help of experienced users. Specialised mounting medium was used to stain nuclei with 4', 6-diamidino-2-phenylindole (DAPI) and protect fluorescent signals from photobleaching (Vectashield DAPI). Cells were prepared as follows:

---

Draw a 1cm diameter circle on a Poly-lysine coated slide using a hydrophobic pen

Drop 200µl of cell suspension into the centre of the circle, Incubate for 20 minutes at 37°C

Wash by briefly dipping into PBS

Add one drop of Vectashield DAPI soft mount (Vector Laboratories) to cells

---

Cover with a coverslip, seal edges with clear nail polish and place slides on ice in the dark until analysis

---

Slides were then examined under a x40 objective using an Axiovert 200M/LSM 510 confocal microscope (Zeiss). Images were viewed and analysed using LSM510 Meta software. Z-stack processing was used to view the distribution of fluorescence within the cell.

### **2.13.7 RNA interference**

RNA interference was performed using a commercially designed pool of 4 sequence specific siRNAs directed against 4 different GATA-1 regions known as ON-TARGET *plus* SMART pool (Dharmacon). A number of experimental controls were also performed. Transfection efficiency was estimated by co-transfection of a non-targeting, non-functional siRNA with a Rhodamine fluorescent tag (siGLO RISC-free siRNA) which can be viewed using fluorescent microscopy or flow cytometry. This siGLO reagent also controlled for off target effects produced by introduction of any siRNA into the cell. A third control was provided by use of a pool of non-targeting but functional siRNAs (siCONTROL) capable of activating RISC and the siRNA pathway but not producing gene specific effects. Experimental design and optimisation are discussed further in Chapter 5, section 5.5.6.

Cells were grown for 48 hours in antibiotic free medium then seeded in 6-well plates at a density of  $5 \times 10^5$  cells per ml (50% confluency). Following initial optimisation experiments transfection was performed using Lipofectamine 2000 reagent (Invitrogen) using the following method:

---

Resuspend siRNA (siCONTROL, siGLO or SMART pool) in 1x siRNA buffer (supplied) to produce a stock concentration of 20 $\mu$ M.

Mix gently and place on orbital shaker for 30 minutes at room temperature

Verify the concentration of siRNA by measuring the A<sub>260</sub>

Dilute 150pmol (7.5 $\mu$ l of 20 $\mu$ M stock) of test siRNA (SMART pool or siCONTROL) and 50pmol (2.5 $\mu$ l) of siGLO in 250 $\mu$ l Opti-MEM I reduced serum medium (Invitrogen), mix gently

Dilute 5 $\mu$ l of Lipofectamine 2000 reagent in 250 $\mu$ l Opti-MEM I. Incubate for 5 minutes at room temperature

Combine the diluted lipofectamine and siRNA, mix gently and incubate for 20 minutes at room temperature then add to the well containing cells and medium and rock gently to mix.

---

Incubate at 37°C for 24-96 hours before assaying for gene knockdown using western blotting or qPCR

---



## **2.14 Embryonic stem cell methods**

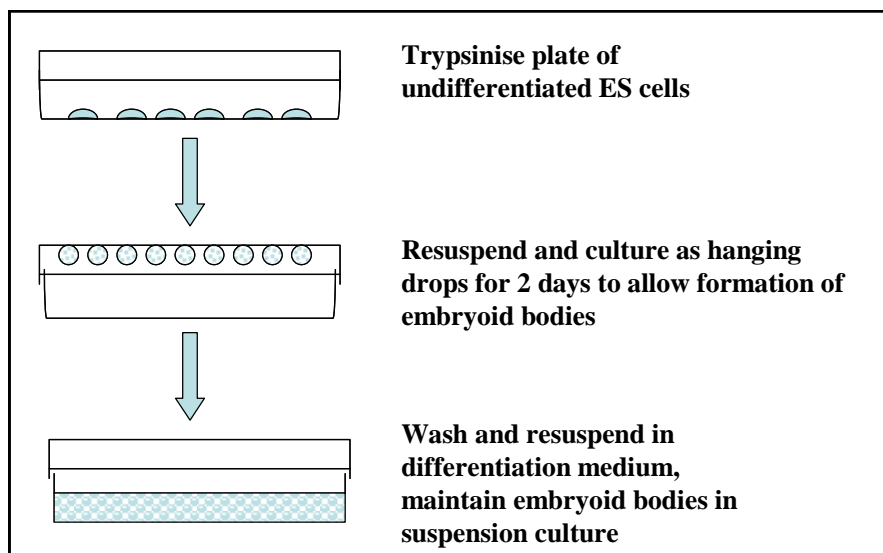
### **2.14.1 *Special considerations for embryonic stem cell cultures***

In order to maintain ES cells in an undifferentiated state a number of additional precautionary steps are needed during routine maintenance of ES cell culture. In particular cells must not be split too thinly, seeded as clumps or allowed to overgrow as all these factors will encourage differentiation. Cells were routinely passaged every 2-3 days, splitting when 70-80% confluent. A single cell suspension was achieved by trypsinisation and manual pipetting. Cells were seeded at  $2 \times 10^5$ /ml on gelatine coated tissue culture flasks (0.1% w/v gelatine in ddH<sub>2</sub>O, autoclaved, 5ml added to plate, incubated for 5-15 mins and then removed). All serum was batch tested by the manufacturer and confirmed as being suitable for ES cell maintenance (StemCell Technologies) and all cells were grown in the presence of recombinant Leukaemia Inhibitory Factor (LIF) 1000units/ml (ESGRO, Chemicon Intl). Cells were inspected daily for evidence of differentiation (colonies surrounded by flattened cells that differ morphologically from central undifferentiated cells- often darker and with long processes and prominent nucleus). Low passage cells were used whenever possible and fresh low passage stocks obtained from the Beatson Laboratories Gene Targeting facility and Dr Lesley Forrester (Scottish Centre for Regenerative Medicine, Edinburgh) for gene targeting experiments.

### **2.14.2 *In vitro differentiation of ES cells***

#### **2.14.2.1 Transcriptional profiling**

ES cells were allowed to differentiate by formation of embryoid bodies followed by withdrawal of LIF and maintenance of embryoid bodies in suspension culture as previously described (Baird, *et al* 2001, Jackson, *et al* 2002). Cells were harvested daily for 7 days, placed in Trizol reagent, RNA extracted and gene expression assessed by PCR. The process is illustrated in fig 2.2 and described below:



**Fig 2.2 in vitro differentiation of ES cells**

Trypsinise one plate of undifferentiated ES cells, wash and resuspend at a concentration of  $2 \times 10^4$  cells/ml

Using a multichannel pipettor dispense 100 drops ( $20 \mu\text{l}$  (400 cells) per drop) concentrically onto the lid of a 10cm tissue culture dish

Invert the lid and place over the bottom of the tissue culture dish filled with 2ml of sterile PBS, incubate at  $37^\circ\text{C}$  to allow formation of cellular aggregates (embryoid bodies)

After 2 days use 5ml of ES cell differentiation medium (no antibiotics, no LIF) to wash the embryoid bodies off the top of the dishes and transfer to a universal container

Spin for 3 minutes at 1200rpm and wash with 5-10ml of ES cell differentiation medium

Resuspend in 10ml differentiation medium and transfer to a 10cm Petri dish (note not a tissue culture dish as this would allow attachment of the embryoid bodies to the plate)

Incubate at  $37^\circ\text{C}$  and harvest one plate daily for 7 days

### 2.14.2.2 Colony formation assays

Colony formation assays were performed by plating ES cells in methylcellulose medium (ES-CULT M3120, StemCell Technologies) in the presence of haematopoietic cytokines. Proprietary 35mm culture dishes recommended for colony forming assays were used (StemCell Technologies). The differentiation medium comprises:

ES cell	ES-CULT (2.3%)	15.6 ml (0.9%)	StemCell technologies
methycellulose	IMDM	19.1 ml	Invitrogen
differentiation	FCS	4 ml (10%)	StemCell technologies
medium (for	Monothioglycerol (1%)	150 $\mu\text{l}$ (340 $\mu\text{M}$ )	Sigma-Aldrich
40ml)	rhuEPO 500U/ml	80 $\mu\text{l}$ (1U/ml)	R&D systems

Insulin (10mg/ml)	40µl (10µg/ml)	Sigma-Aldrich
rmuIL-3 (100µg/ml)	16µl	R&D systems
L-glutamine 100x	400µl	Invitrogen
Pen-Strep 100x	200µl	Invitrogen

Method for performing in vitro differentiation assay:

Thaw methylcellulose medium for 24 hours at 4°C prior to addition of cytokines as described above, aliquot 1.5ml of this methylcellulose mix per 35mm tissue culture plate using a 16G blunt ended needle and syringe, ensuring no air bubbles are formed

Grow undifferentiated ES cells on gelatine with LIF, for at least one passage after thawing, change medium a few hours before trypsinising.

Wash cells with PBS, trypsinise and resuspend at  $1 \times 10^5$  cells per ml in IMDM/10%FCS

Take up 12.5µl (1250 cells) of ES cell suspension in a pipette and eject into the aliquoted methylcellulose medium with a swirling motion to ensure even distribution of the suspension in the semi-solid media

Place 2 plates into a 100mm Petri dish with a third 35mm dish containing 3-4ml sterile water to provide humidity. Incubate at 37°C for up to 14 days observing plates daily from day 5 onwards for formation of haematopoietic colonies

Qualitative and quantitative analysis of colonies on days 10-14 using a counting grid, BFU-E identified on the basis of morphology and pale pink/red haemoglobinisation, CFU-GM on the basis of characteristic morphology.

### ***2.14.3 Picking and storing targeted ES cell clones for screening***

10-14 days following electroporation of the targeting vector and growth in selective medium ES cell colonies were picked for screening for the targeted allele. Duplicate plates were made for each clone – one master plate to be frozen down and a test plate for genomic DNA extraction. The following method was used:

Add 20µl trypsin to each well of a 96-well plate. Add 500µl of ES cell media to each well of 8 x 24-well plates (96 wells in duplicate). Label 24-well plates A1-12, B1-12 etc in duplicate (master and test), incubate at 37°C until use.

Wash ES cell colonies with warmed PBS (37°C), leave about 2-3ml PBS in the plate to prevent drying

Using a p200 Gilson pipette set to aspirate 150µl, draw up a small amount of trypsin from well A1, use this tip to scrape/aspirate a colony then eject the tip back into the well A1.

---

Repeat for wells A2-A12.

Reattach the tip from well A1 to the Gilson pipette, draw up 150µl of media from the corresponding well in the 24-well master plate, return to well A1, pipette up and down 3 times and then draw up total volume (150µl) and split evenly between corresponding duplicate wells in the master and test 24-well plates.

Repeat for wells A2-A12

Repeat the process for the next 12 colonies B1-B12 etc until all the colonies are picked

---

Incubate master and test plates at 37°C until the majority of wells are confluent. At this stage freeze down the master plate (see section 2.14.3.1 below) and extract genomic DNA from the test plate (section 2.7.1)

---

### **2.14.3.1 Freezing and Thawing 24-well master plates**

---

To freeze a 24-well master plate:

---

Aspirate the medium, wash in PBS x1

Add 100µl trypsin 0.05% to each well, incubate 37°C x 5 minutes

Add 200µl of 1.5x freezing medium (8.5ml ES cell media, 1.5ml DMSO) and pipette up and down to mix

Seal edges of plate with tape, place on ice, wrap in cotton wool and place in polystyrene box at -80°C

---

---

To thaw a 24 well master plate:

---

Float plate in a 37°C waterbath

Once thawed transfer contents of each well to a fresh 24-well plate containing 0.5ml fresh media, incubate at 37°C

Once cells have adhered to the wells (approximately 8 hours later), remove DMSO containing media and replace with fresh warmed ES cell media.

---

## **2.15 Statistical methods**

Statistical analysis was performed using Graph Pad Prism software. A two-tailed student's t-test was used to compare normally distributed variables between two test groups. Analysis of variables in three test groups at more than one time point was performed using two-way ANOVA analysis. In all cases a p value of  $\leq 0.05$  was considered significant.

### 3 An analysis of endogenous GATA-1 isoform expression in primary tissues and cell lines

---

The GATA-1s isoform was first reported in 1995 (Calligaris, *et al* 1995). This paper described the presence of a 40kDa GATA-1 band in addition to the expected 47kDa full-length band on western blotting of mouse and human haematopoietic cell lines. This was shown to be a shortened form of GATA-1 lacking the N terminus. Expression of the short form was documented in early embryonic development and was said to precede expression of the FL form. However the published western blot is difficult to interpret (reproduced in Fig 3.10b) and this finding has not been independently confirmed. Functional assays in heterologous cell lines revealed a reduced transactivation potential for GATA-1s compared to GATA-1FL (Calligaris, *et al* 1995). This role of the N-terminus in enhancing transactivation of genes carrying upstream GATA-1 binding sites, was confirmed by experiments looking at the functional domains of GATA-1 in primitive and definitive haematopoiesis (Shimizu, *et al* 2001). Little progress was made in elucidating the biological role of GATA-1 isoforms until the description of GATA-1 mutations in DS-AMKL (Wechsler, *et al* 2002). This paper rekindled interest in the physiological and pathological roles of GATA-1s.

In order to contribute to an investigation of the possible functions of GATA-1s in haematopoiesis, the expression pattern of the two isoforms during ontogeny and in various primary tissues and cell lines was investigated. As discussed in 1.2.8.1 the production of GATA-1s in mice is thought to arise by alternative translation of a single mRNA (Calligaris, *et al* 1995), whereas more recent evidence has emerged for alternative splicing in humans (Rainis, *et al* 2003). Whether mice also produce alternatively spliced transcripts is unknown. This chapter describes experiments to determine:

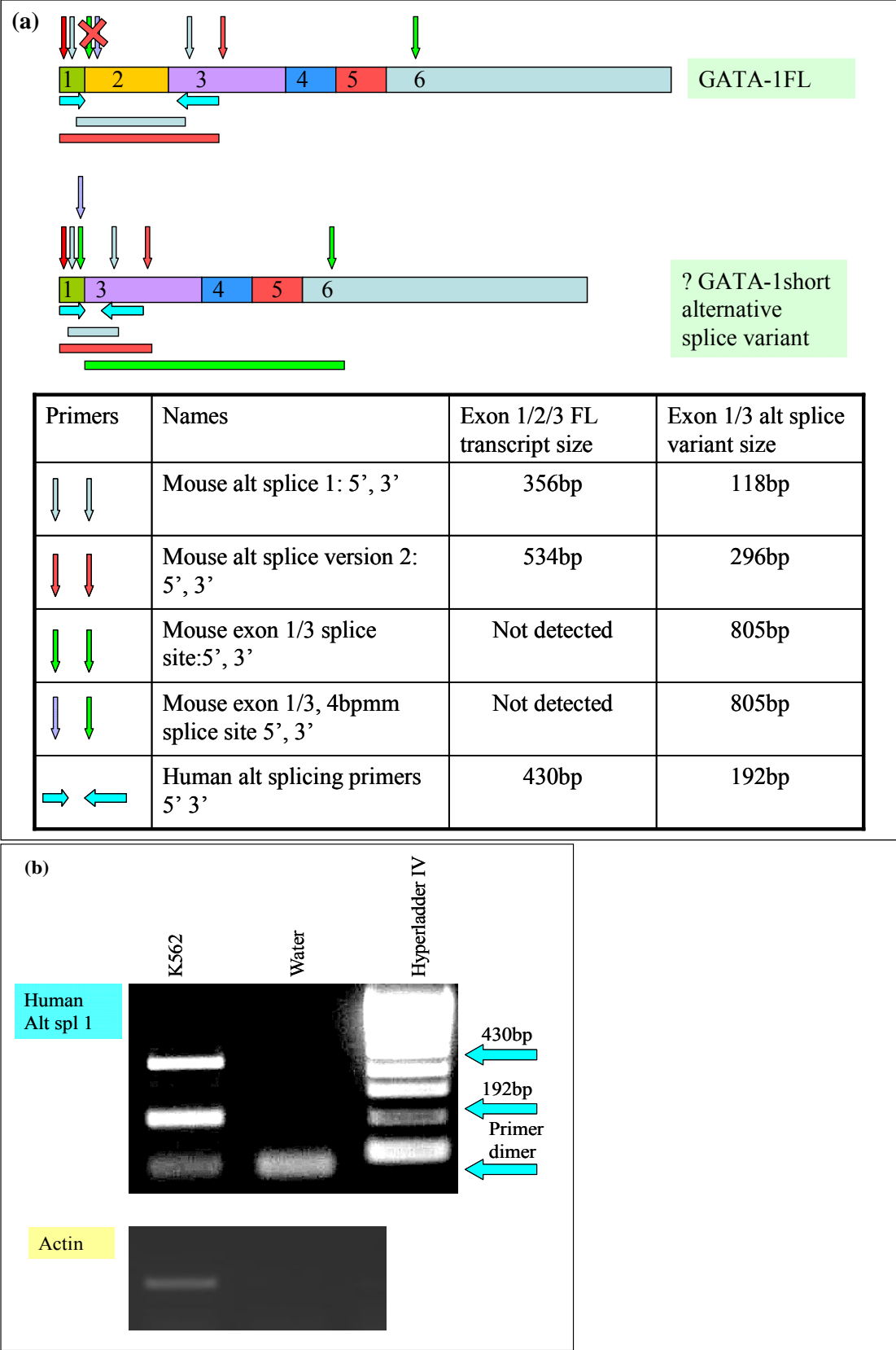
1. The best methods to detect isoform expression in humans and mice and whether alternative splicing also plays a role in the murine production of GATA-1s.
2. Patterns of GATA-1FL and GATA-1s expression during ontogeny, in different cell lineages and during in vitro haematopoietic differentiation.
3. Whether additional GATA-1 isoforms exist.
4. The relevance of an identified human leukaemic transcript variant BC009797.

In this way it is hoped to determine whether GATA-1s has a distinct pattern of expression which could shed light on its biological role in vivo. In addition, in order to understand the complex roles of GATA-1 in haematopoiesis it is vital to identify any additional isoforms which could play novel roles in GATA-1 regulation of gene expression.

### **3.1 GATA-1 isoforms are produced by different mechanisms in humans and mice**

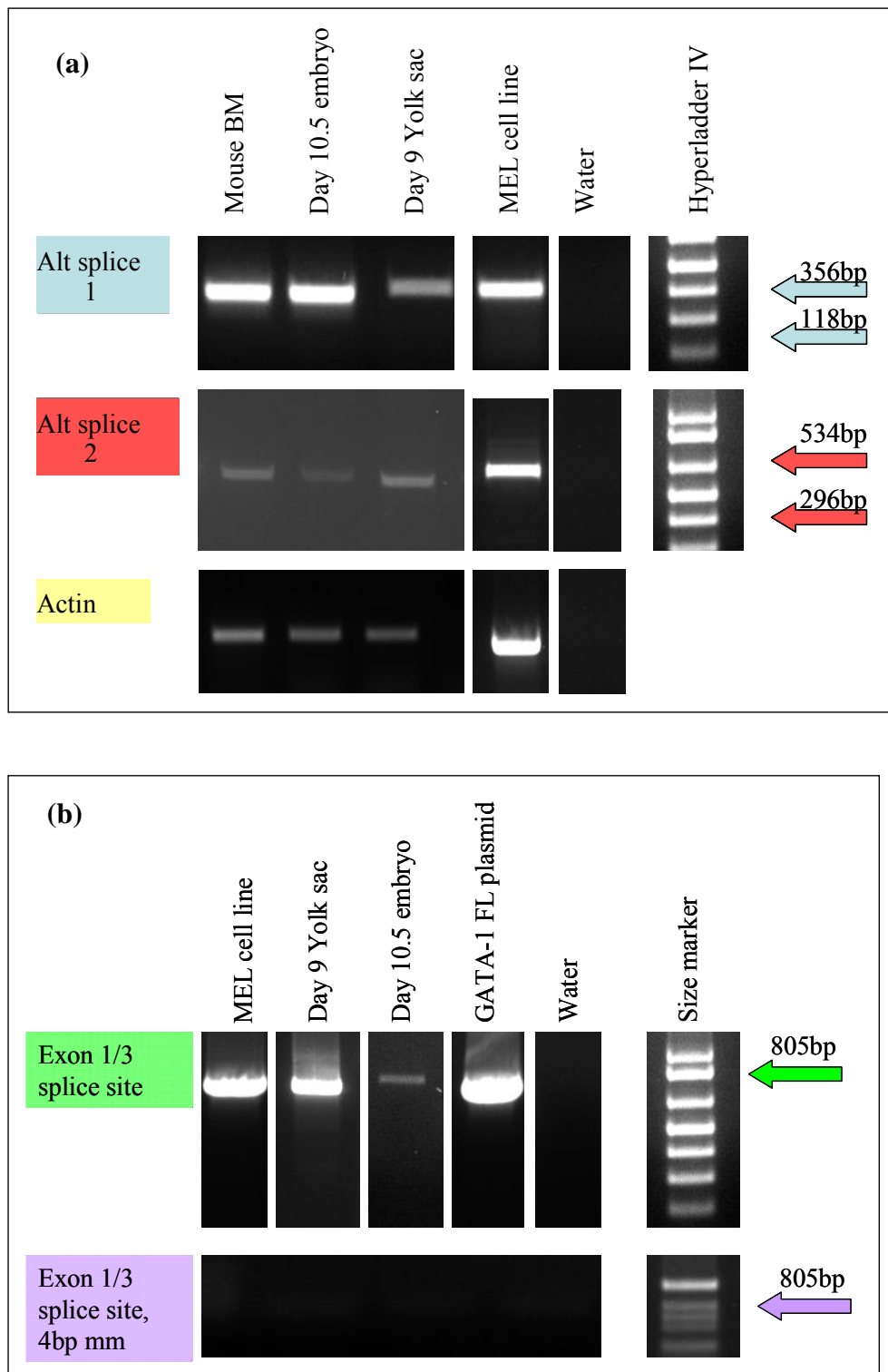
#### ***3.1.1 PCR primers spanning potential splice sites show one product in mice and two products in humans.***

To investigate the existence of alternative splicing in mice, and confirm its occurrence in humans, a number of PCR primers were devised. These are illustrated in Fig 3.1a and listed in table 2.4. In the presence of alternative splicing, primers spanning the splice sites should produce two different sized products; one corresponding to the exon 1-2-3 splice variant and the other corresponding to the exon 1-3 splice variant which should be 238bp (human and mouse) smaller due to the skipping of exon 2. Initial results using these primers confirmed the presence of two bands in cDNA from the human cell line K562 (fig 3.1b). These bands were excised, gel purified and cloned into the pCR2.1 TOPO cloning vector (section 2.12.1.1) before being sent for commercial sequencing (Agowa.de). Sequencing results confirmed that the larger band corresponded to the exon 1-2-3 splice variant and the shorter band lacked exon 2 and showed direct splicing from exon 1 to exon 3 (data not shown). Only one band was seen in mouse bone marrow cDNA (fig 3.2a). Since GATA-1s is thought to be produced early in ontogeny (Calligaris, *et al* 1995), and may not be needed for adult haematopoiesis (Shimizu, *et al* 2001), it is possible that this result reflected lack of production of the GATA-1s isoform in adult tissues, rather than lack of alternative splicing. Therefore, murine samples from embryonic tissues (yolk sac day 9, embryo day 10.5) were also examined. Again only one band was seen (fig 3.2a). Finally a murine cell line - mouse erythroleukaemia (MEL) - that expresses both isoforms by western blot (fig 3.3b) was tested, again despite the existence of long and short isoforms at the protein level there was no evidence of alternative splicing at the mRNA level (fig 3.2a). This result was confirmed using an independent set of primers (alt splice 2, fig 3.2a).



**Fig 3.1 Detection of alternative splicing by PCR**

(a) Details of primer design and expected product sizes, exon 1/3 primers should not anneal to the GATA-1FL transcript as shown by a red cross (b) clear evidence of alternative splicing in the human cell line K562, results were confirmed by cloning and sequencing.



**Fig 3.2 PCR does not reveal alternative splicing in mice**

**(a)** 2 different primer sets only detect a single GATA-1 transcript. **(b)** Primers designed to anneal to the putative exon 1/3 splice site are still able to detect the full length transcript unless a 4bp mismatch is introduced. This mismatch prevents non-specific annealing but is unable to detect a short transcript. Primer binding sites and products are shown in Fig 3.1a.

To increase the sensitivity of the technique, in an attempt to detect a low level of exon 1/3 transcripts, PCR primers were designed that annealed to the theoretical exon 1/3 splice junction (fig 3.1a). Unfortunately due to a high degree of homology between the exon 1/3 junctional region and the 1/2 junctional region these primers were able to pick up the full-



length transcript (fig 3.2b, top gel image), to overcome this a 4 base-pair mismatch was introduced into the primer which should leave sufficient homology to pick up the exon 1/3 splice junction but should abolish binding to the already mismatched exon 1/2/3 junctional site. The latter was indeed confirmed (fig 3.2b, bottom gel image) but this primer was still unable to detect an alternatively spliced transcript.

To strengthen these findings the mouse expressed sequence tag (EST) database (see 3.3.1.2 for a discussion of this methodology) was examined for evidence of published mRNA transcripts corresponding to an exon 1/3 splice variant using the National Center for Biotechnology Information (NCBI) Basic Local Assignment Search Tool (BLAST) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). This in silico analysis also failed to confirm the presence of alternative splicing in mice (data not shown). A similar search of the human EST database showed evidence of an Exon 1/3 alternatively spliced transcript in a cDNA library constructed from human fetal skeletal muscle (NCBI accession number: AJ973498).

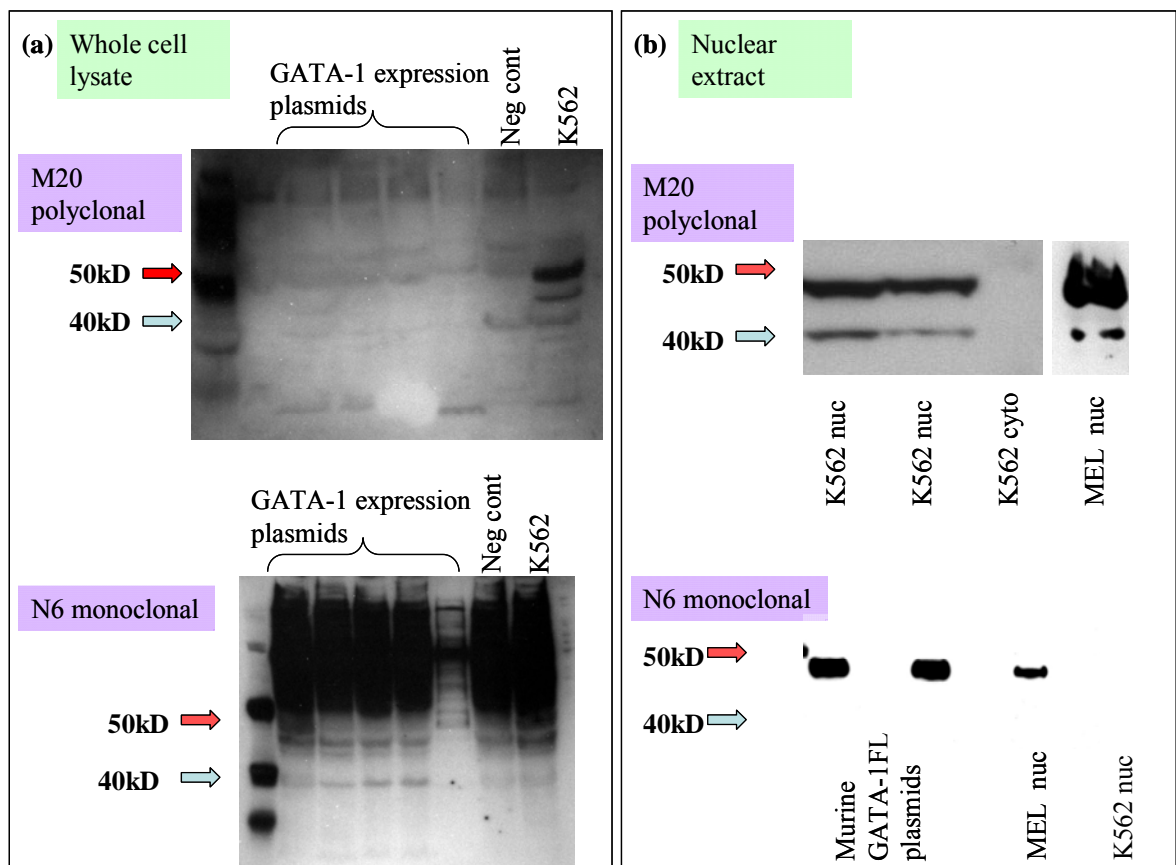
*In summary, these results confirm the existence of alternative splicing as a method of short isoform production in humans. They also strongly support the hypothesis that the short isoform is produced by alternative translation, rather than alternative splicing, in mice. These findings have implications for investigation of isoform expression patterns as PCR will only be informative in human samples, alternative methods of detection will be needed for mice.*

### **3.1.2 Western Blotting confirms presence of 2 isoforms in mice and humans**

In the absence of alternative splicing, detection of GATA-1 isoform production in mice relies upon Western blotting. There are two commercially available anti-mouse GATA-1 antibodies. A polyclonal antibody raised against the conserved C-terminus GATA1-M20 (Santa Cruz) and a monoclonal antibody raised against the N-terminus – GATA1-N6 (Santa Cruz). Because GATA-1s lacks the N terminus it is predicted that it will not be detected by the N6 antibody but should be detectable as a smaller band of approximately 40kDa when using the polyclonal M20 antibody. GATA-1FL should be detectable as a 47kDa band with both antibodies. These antibodies are reported as suitable for detection of both human and murine proteins on the product datasheet.

### 3.1.2.1 Optimisation of western blotting techniques

Initial attempts at western blotting followed standard protocols for whole cell lysates (see materials and methods section 2.9). These blots showed a large amount of non-specific binding. Differentiation between background and GATA-1 was impossible (Fig 3.3a). This problem persisted despite titration of primary and secondary antibody concentrations, varying incubation times, addition of bovine serum albumin, trying alternative lysis agents and increasing the stringency of washing.



**Fig 3.3 Optimisation of western blotting technique**

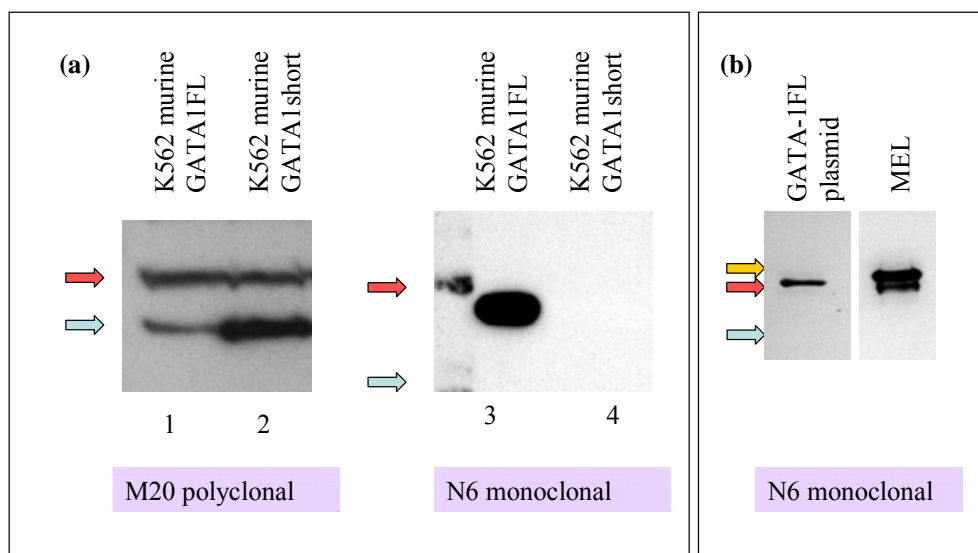
(a) Whole cell extracts probed with M20 and N6 antibodies showing high level background and non-specific bands in negative control (b) The use of the same antibodies and blotting conditions but with nuclear extracts. The 47kDa GATA-1FL band is seen in with both M20 and N6 antibodies whilst the smaller 40kDa GATA-1s band is only seen with M20. Red arrow=50kDa and blue=40kDa size markers respectively. GATA-1 was not detectable in cytoplasmic extracts from the same cells (lane 3 M20).

Since GATA-1 is a nuclear transcription factor a nuclear extraction protocol was attempted to concentrate the protein of interest and reduce non-specific background. This approach was much more successful as shown in Fig 3.3b. Using this approach the M20 antibody produced two bands corresponding to GATA-1FL (47kDa) and GATA-1s (40kDa) in both human (K562) and murine (MEL) cell line nuclear extracts (Fig 3.3b, top image). As

predicted the N6 antibody only produced a 47kDa band corresponding to GATA-1FL (Fig 3.3b, bottom image), this band is shown in NIH3T3 cells expressing two murine GATA-FL plasmids (described in chapter 5.2.2) and in a murine erythroleukaemia cell line (MEL). Interestingly, the N6 antibody did not produce any bands with a human K562 nuclear extract (lane 4 fig 3.3b bottom image) despite the clear detection of a GATA-1FL band using the M20 antibody (Fig 3.3b, top image). This finding is discussed in section 3.1.2.2 below. The nuclear/cytoplasmic extraction protocol (described in section 2.9.2) was used in all subsequent blotting experiments. In order to ensure cytoplasmic expression was not being missed, cytoplasmic extracts were run on a number of occasions – these never showed bands corresponding to GATA-1FL or GATA-1s (fig 3.3b, top image lane 3).

### 3.1.2.2 GATA-1 N6 antibody only detects murine GATA-1

Although product information claims that the N6 monoclonal antibody picks up both murine and human GATA-1, results shown above suggest that only murine protein can be detected. To test this further a number of transfection experiments using GATA-1 expression plasmids (described in chapter 5.2.2) were performed. Fig 3.4a shows a western blot of a mouse GATA-1FL plasmid transfected into a human cell line (K562) alongside a mouse GATA-1s plasmid transfected into the same cell line. The M20 antibody picks up both human and mouse isoforms (lane 1 and 2) but the N6 antibody is only able to detect the mouse GATA-FL transcript (lane 3) and not the human one (lane 4).



**Fig 3.4 Characteristics of available antibodies**

(a) M20 antibody detects murine and human GATA-1 whilst N6 only detects murine protein  
 (b) A double band was sometimes detected in cell lines expressing endogenous GATA-1.  
 Red arrow – position of GATA-1FL, blue arrow- position of GATA-1s, gold arrow – additional band.

This finding led to the use of the M20 antibody (able to detect both isoforms in both species) in the majority of experiments.

### **3.1.2.3 An intermittent additional band suggests variable post-translational modification**

As shown in Fig 3.4b an additional band could sometimes be detected just above the FL isoform (fig 3.4b). This double band was not consistently detected and was only ever seen when using cell lines. On review of the literature this pattern has been reported elsewhere with the larger band being identified as a phosphorylated form of GATA-1 (Crossley and Orkin 1994, Hernandez-Hernandez, *et al* 2006).

*In summary, these experiments confirm the existence of two isoforms of GATA-1 at the protein level in humans and mice. Methods of isoform detection by western blotting have been optimised, and important differences in antibody species reactivity identified.*

## **3.2 Patterns of GATA-1 isoform expression during haematopoietic development**

The studies described above establish that GATA-1 isoforms can be identified by PCR in humans and by western blotting in mice. Although the sites and timing of GATA-1 expression have been extensively studied during haematopoietic ontogeny, very little is known about the relative utilisation of the different isoforms during this period. Apart from the intriguing report of GATA-1s expression preceding that of GATA-1FL in the early mouse embryo (Calligaris, *et al* 1995), there have been no other studies addressing differential isoform usage in different lineages or at different developmental stages either in humans or mice. Therefore, a systematic analysis of the sites and timing of isoform expression in humans and mice was undertaken using the following material:

1. Human primary haematopoietic cells
2. Human haematopoietic cell lines
3. Human embryonic tissues
4. Human haematopoietic cell lines undergoing in vitro megakaryocytic and erythroid differentiation

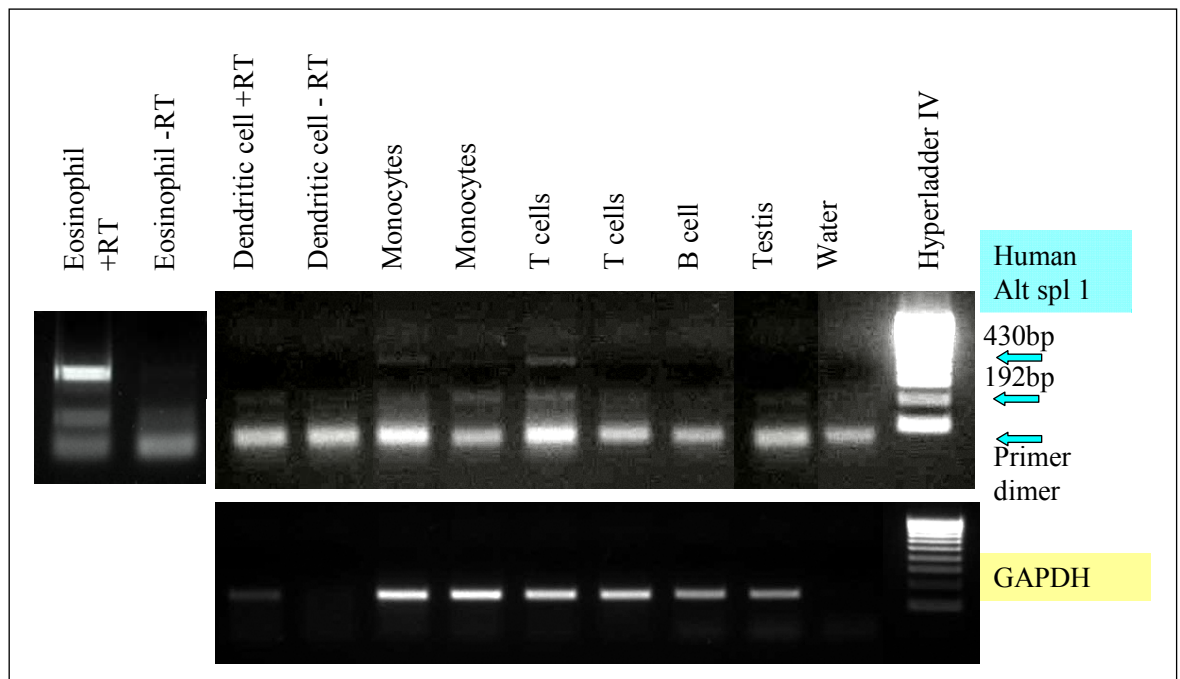
5. Murine embryonic tissues

6. Murine cell lines

7. Murine embryonic stem cells undergoing in vitro haematopoietic differentiation.

### **3.2.1 GATA-1 isoform expression in human primary cells**

GATA-1 is known to be expressed in the erythroid, megakaryocyte, eosinophil, mast cell and dendritic cell lineages as well as the Sertoli cells of the testes (Crispino 2005), however isoform expression patterns in these cell types have not been reported. The existence of GATA-1s expression would provide circumstantial evidence for a functional role in these lineages. Expression patterns were analysed in various haematopoietic primary cells and in adult testis. Purified monocyte (2 preparations), eosinophil, T cell (2 preparations), B cell and dendritic cell populations were obtained from Dr C McKimmie, Glasgow Biomedical Research Centre. Monocytes, eosinophils and lymphocytes were isolated by a standard MACS bead sorting approach (McKimmie, *et al* 2008) and dendritic cells isolated following Flt3 stimulation of in vitro cultures. RNA extraction using a commercial kit (RNeasy, Qiagen) and cDNA synthesis using AffinityScript (Stratagene) was performed as described in section 2.8.1.2 and 2.8.4. Testis cDNA was obtained commercially (Ambion). Expression was assessed by RT-PCR using exon 1 and 3 primers which should give a 430bp band with GATA-1FL and a 192bp band with GATA-1s. Results are shown in Fig 3.5. Both isoforms were expressed in eosinophils. Low level GATA-1FL and GATA-1s expression was seen in both monocyte preparations with one sample appearing to show a stronger GATA-1s band, one T cell sample was also positive but dendritic cells, B cells and the other T cell population were all negative. Although GATA-1 expression has been documented in dendritic cells previously (Gutierrez, *et al* 2007) this was in mice not humans. In addition, the intensity of GATA-1 expression in murine DCs varies with their activation status with very low levels of expression in steady state DCs and significant up-regulation on activation and terminal differentiation (Gutierrez, *et al* 2007). It is possible that dendritic cells used in this experiment were not mature enough to express detectable GATA-1, alternatively the relatively small amounts of template (as shown by the weak GAPDH band) may have led to GATA-1 being below the level of detection in this PCR reaction. The expression in monocytes is interesting as these cells can act as precursors for dendritic cells under inflammatory conditions but their other cellular progeny - macrophages are not known to express GATA-1. Expression was not detected in testis consistent with use of an alternative first exon (Ito, *et al* 1993).



**Fig 3.5 GATA-1 isoform expression in human primary cells**

PCR was performed using exon 1 and 3 primers. Both isoforms are expressed in eosinophils, and T cells, monocytes also appear to express GATA-1 with one sample (lane6) expressing a stronger GATA-1s band.

These experiments, using purified primary cells, always bear the caveat that low levels of contamination with other populations may be responsible for the detection of expression – especially when using sensitive techniques such as PCR. To explore these cell types further, without concern about contamination, the expression of GATA-1 isoforms in a number of cell lines was also studied.

### **3.2.2 GATA-1 expression patterns in human cell lines**

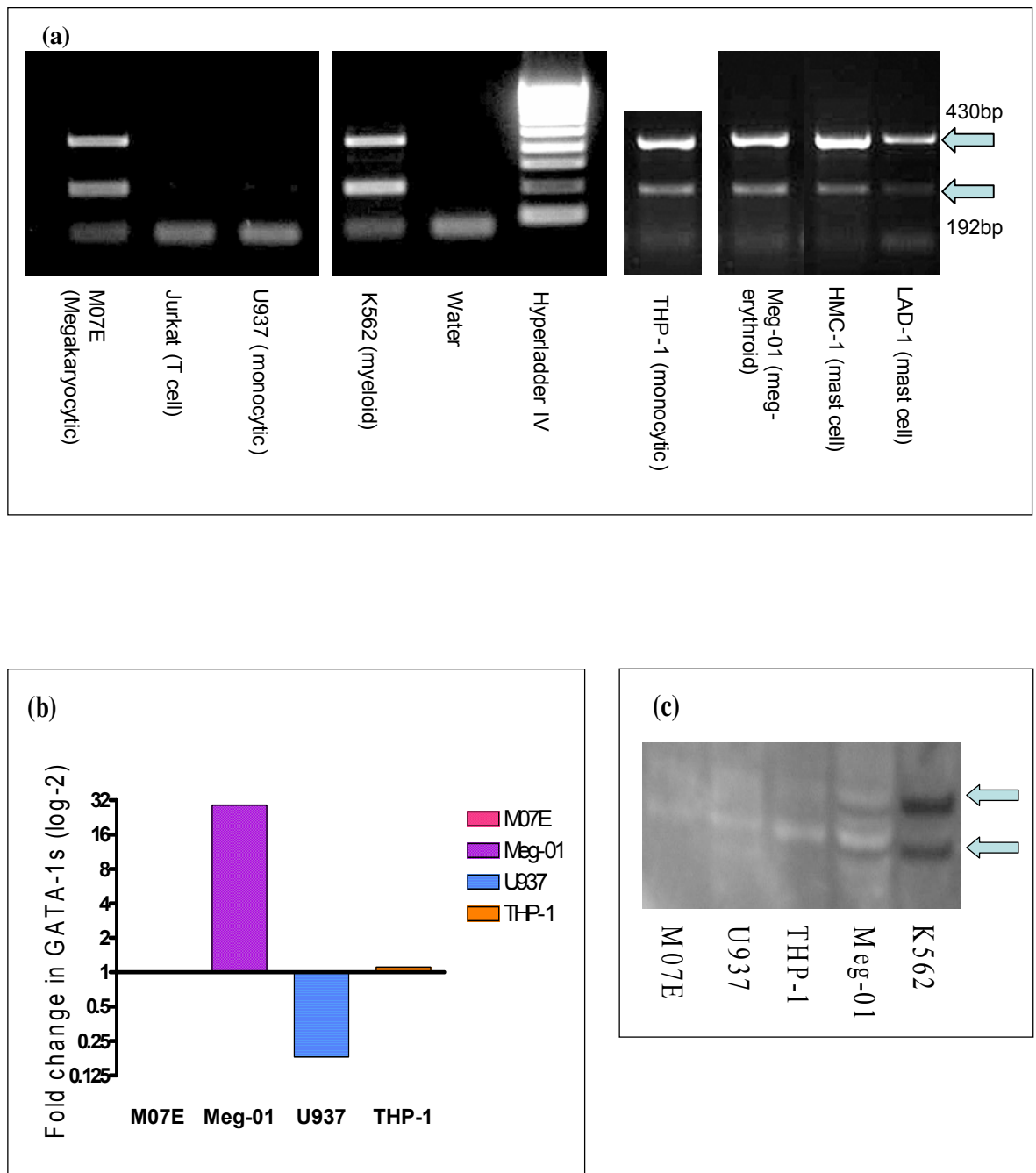
Cell lines have the advantage that they can be grown in large quantities and represent pure populations without concerns about cross-contamination. However, they may not be entirely representative of the cell type from which they were derived and have acquired mutations to allow them to proliferate indefinitely in culture. In addition many are derived from malignant cells whose differentiation status and expression of lineage specific genes may be considerably disturbed. These factors should be borne in mind when interpreting experiments.

A number of human haematopoietic cell lines were grown in culture as described in section 2.2 and 2.5.3. These are listed in Table 3.1 along with their presumptive cell of origin.

Cell line	Putative cell of origin
M07E	Megakaryocyte precursor
Jurkat	T cell
U937	Monocyte
THP-1	Monocyte
K562	Common myeloid precursor
Meg-01	Meg-erythroid precursor
HMC-1	Mast cell
LAD-1	Mast cell

**Table 3.1 Human haematopoietic cell lines and their normal counterparts**

Once sufficient cells had been obtained these cultures were harvested and RNA extraction and nuclear protein extraction were performed according to standard protocols (described in sections 2.8.1. and 2.9.2). cDNA was made from the RNA using AffinityScript with random primers (section 2.8.4) and used in RT-PCR with exon 1 and 3 primers (Fig 3.1). Results of PCR reactions are shown in Fig 3.6a. Expression of both isoforms was seen in the megakaryocytic cell lines Meg-01 and M07-E along with K562 (a cell line known to be capable of meg-erythroid differentiation as shown in section 3.2.4 below). In addition the mast cell lines HMC-1 and LAD showed strong expression levels of both isoforms. Mast cells are known to require GATA-1 for terminal differentiation (Migliaccio, *et al* 2003). These results also confirmed GATA-1FL and GATA-1s expression in one of two monocytic cell lines – THP-1. The T cell line Jurkat did not show any expression. To quantitate levels of expression of GATA-1s in megakaryocytic and monocytic cells qPCR was performed as described in section 2.10.5 – M07E and THP-1 cells showed equivalent, albeit low level, GATA-1s expression, the other monocytic cell line U937 showed negligible expression, whilst high level expression was confirmed in Meg-01 cells (fig 3.6b). Western blotting (fig 3.6c) failed to show GATA-1 expression in both M07E and THP-1 presumably reflecting the low level expression in these cells. K562 and Meg-01 cells expressed both isoforms at the protein level as expected.



**Fig 3.6 GATA-1 expression in human cell lines**

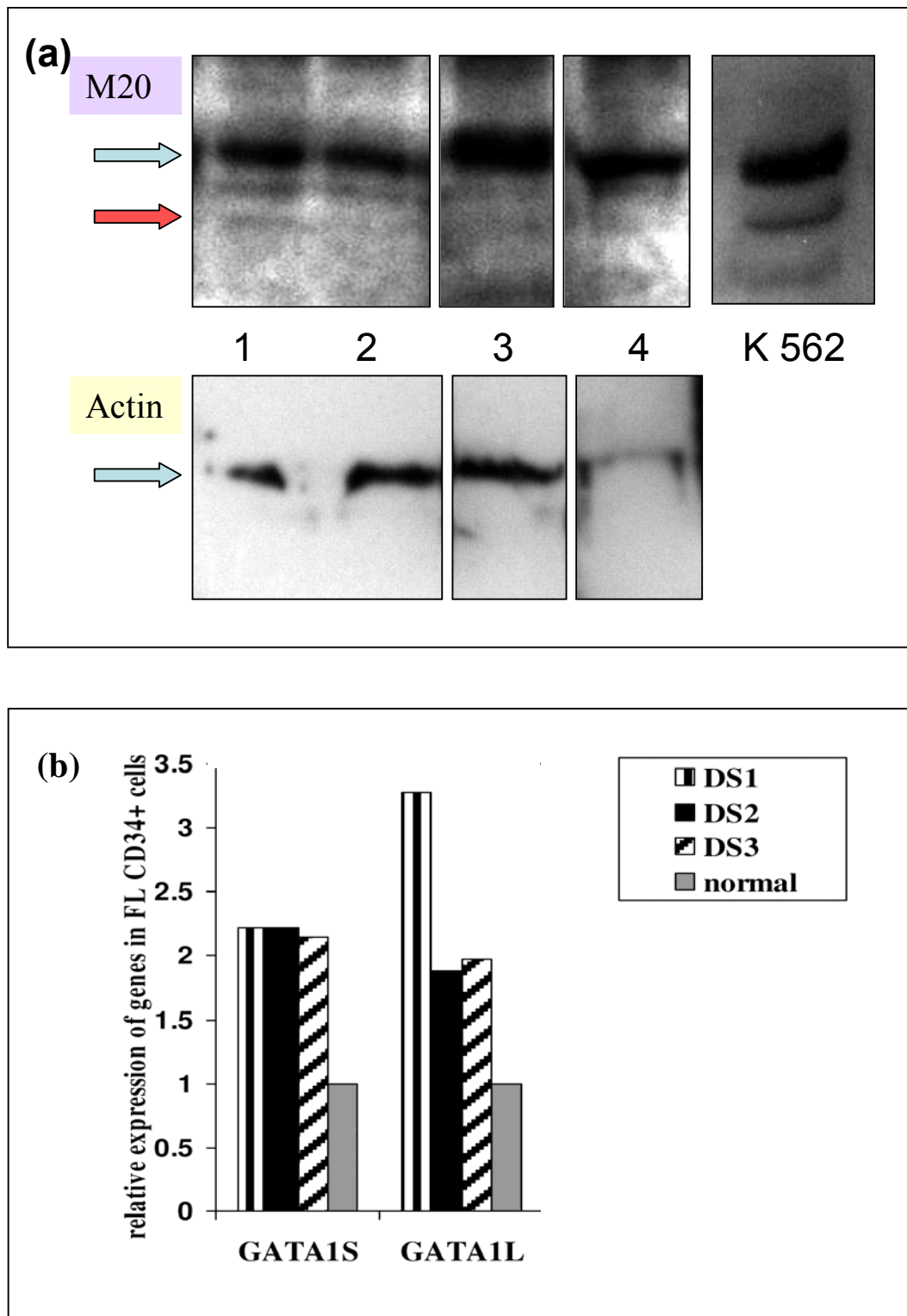
(a) Expression of isoforms assessed by PCR confirms expression in meg-erythroid cell lines (M07E, Meg-01, K562), mast cells (HMC-1, LAD-1) and 1 monocytic cell line (THP-1). Arrows show position of the 430bp GATA-1FL band and the 192bp GATA-1s band. (b) qPCR for GATA-1s using levels of expression in M07E cells as the calibrator (i.e. level of expression in M07E arbitrarily set at 1) shows equivalent levels of expression in THP-1 and increased expression in Meg-01 ( $n=3$ ). (c) western blot analysis (M20 antibody) is unable to detect low level THP-1 and M07E GATA-1s expression but confirms 2 isoforms in Meg01 and K562, top arrow shows position of GATA-1FL and bottom arrow position of GATA-1s.



*In summary, studies looking at GATA-1 isoform expression in human primary cells and haematopoietic cell lines have shown that both isoforms are expressed in megakaryocytic, erythroid, eosinophil and mast cell lineages. Monocytes show variable expression, confirmed by quantitative PCR. Unexpectedly, one primary T cell sample showed evidence of GATA-1 expression although this was not confirmed on examination of a T cell line, suggesting that the detectable expression may have been due to contamination of the preparation with other GATA-1 expressing cell populations. GATA-1s expression appears to accompany GATA-1FL in all the cell types studied, although one monocyte preparation showed preferential GATA-1s expression.*

### **3.2.3 GATA-1 isoform expression during human fetal development**

As discussed in chapter 1 the GATA-1s isoform may be particularly important for fetal haematopoiesis. Human fetal samples are obviously a scarce resource, making comprehensive analysis of isoform expression during human haematopoietic ontogeny difficult. However, in collaboration with Professor Irene Roberts and Dr Oliver Tunstall-Pedoe (Imperial College, London), a second trimester human fetal liver sample derived from elective surgical termination of pregnancy and two neonatal cord blood samples, one of which had undergone in vitro culture to enrich for neonatal megakaryocytes were evaluated. Whole cell lysates were made from provided cell pellets (as described in 2.9.1) and the results of western blotting are shown in fig 3.7. Although not conclusive, these blots suggest that both GATA-1 isoforms be detected in both cord blood samples (lanes 1-3) and in fetal liver (lane 4). The short isoform is most strongly expressed in cells derived from CD61 selected megakaryocytic cord blood cultures (lane 1), with reduced expression of this isoform in the CD61 negative fraction, consistent with a role for GATA-1s in fetal megakaryocytic development. These western blotting results are supported by qPCR for the GATA-1 isoforms (fig 3.7b– analysis performed by Dr O Tunstall-Pedoe). These qPCR results show detectable GATA-1s and GATA-1FL expression in CD34+ cells derived from DS and normal fetal liver samples with a median gestation of 15 weeks, Down syndrome fetal liver samples had upregulated GATA-1 expression consistent with their increased proportion of MEPs, no GATA-1 mutations were detectable at this time.



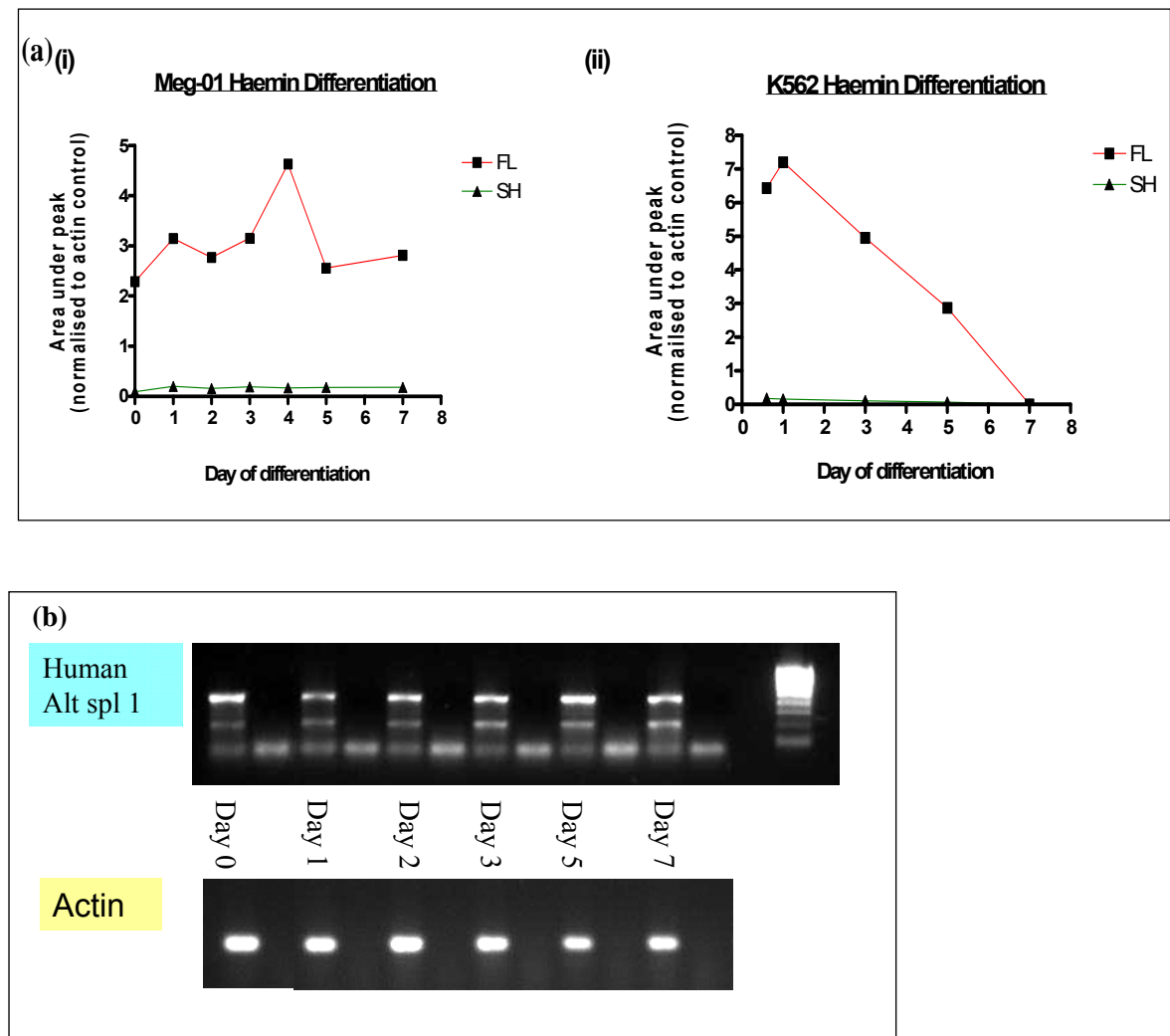
**Fig 3.7 GATA-1 isoform expression in human embryonic tissues**

(a) Western blotting with M20 (top, blue arrow – position of GATA-1FL band, red arrow position of GATA-1s band) and actin loading control (bottom, blue arrow- position of 42kD actin band). Lane 1: CD61 bead selected cord blood (CB) cells following in vitro culture. Lane 2: CD61 negative CB cells from same culture. Lane 3: unmanipulated mononuclear cells from term CB sample. Lane 4: Down syndrome fetal liver sample 15 weeks gestation. (b) qPCR results for GATA-1 isoforms in Down syndrome (DS1, DS2, DS3) and normal fetal liver (FL) samples (normal) from average 15 weeks gestation showing upregulation of both isoforms in the presence of trisomy 21. GATA1S = short isoform, GATA1L = full-length isoform - this qPCR data was reproduced from (Tunstall-Pedoe, *et al* 2008) with permission.

### ***3.2.4 GATA-1 isoform expression during induction of terminal erythroid and megakaryocytic differentiation***

Studies described above addressed the presence or absence of GATA-1 expression in static cell populations. GATA-1 levels are known to be dynamically regulated during erythroid and megakaryocytic differentiation but whether this regulation affects both isoforms equally is unknown. Certain haematopoietic cell lines can be induced to differentiate into erythroid or megakaryocytic cells *vitro*. The chemical compound Haemin ( $C_{34}H_{32}N_4O_4FeCl$ ) induces erythroid differentiation as measured by upregulation of erythroid genes and characteristic cytology. The phorbol ester TPA (12-O-tetradecanoylphorbol-13-acetate) induces megakaryocytic differentiation via a mechanism that is poorly understood but appears to involve activation of protein kinase C. Two cell lines were chosen for study- both originated from patients with CML - K562 and Meg-01. These lines are discussed in more detail in Chapter 5 and data validating this model system are also presented in Chapter 5. Importantly, Meg-01 cells are trisomic for chromosome 21 whilst K562 are disomic, allowing additional evaluation of the role of trisomy 21 in modulating the effects of GATA-1s expression. By serial harvesting of RNA the expression levels of genes during lineage commitment and differentiation can be studied. *In vitro* differentiation was performed (as described in section 2.13.3. and 2.13.4) to test whether GATA-1 isoforms vary during commitment to, and differentiation within, the megakaryocytic and erythroid lineages.

Fig 3.8 shows results for Haemin-induced erythroid differentiation, with assessment using semi-quantitative RT-PCR (see section 2.10.2 materials and methods), for both isoforms. These results suggest that GATA-1FL is dynamically regulated during erythroid differentiation. K562 cells show rapid and profound down-regulation of GATA-1FL from day 3 onwards compatible with the need for GATA-1 down-regulation to allow terminal differentiation –as discussed in chapter 1. Meg-01 cells appear initially to upregulate GATA-1FL before levels fall again at day 4-5. This may represent the need to reverse their partially committed phenotype (they have strong megakaryocytic features in culture) prior to terminal erythroid differentiation. However, the nature of these data (n=1 and use of semi-quantitative PCR) means that caution should be exercised in their interpretation. The molecular basis for erythroid and megakaryocytic differentiation in these cells is further explored by qPCR in chapter 5. For the purposes of this experiment the important observation is that, unlike GATA-1FL, GATA-1s does not show any appreciable change in expression levels with erythroid differentiation.

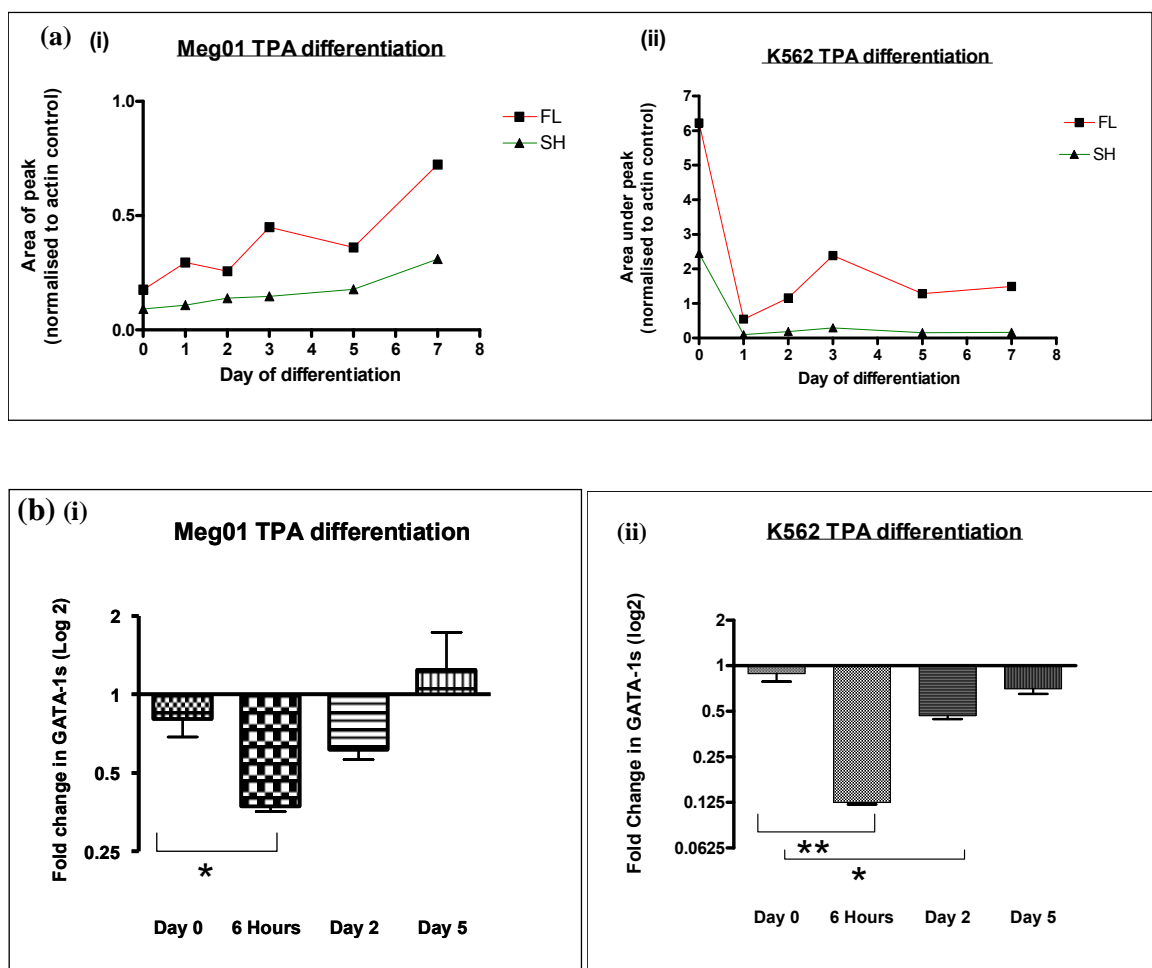


**Fig 3.8 Variation of GATA-1 isoforms during in vitro erythroid differentiation**

(a) Results of densitometry assessment of RT-PCR gels, 25 cycles of PCR were performed for GATA-1 and the actin control, to ensure the reactions were still in the exponential doubling phase. The area under the peak was normalised to the actin control for each isoform ( $n=1$  for each isoform/cell line). (b) Representative 25 cycle gel image showing variation in the intensity of GATA-1 bands and the need to normalise to actin levels. FL = GATA-1FL stable transfectants, SH = GATA-1s stable transfectants.

Fig 3.9 shows results for TPA induced megakaryocytic differentiation. In addition to semi-quantitative PCR, expression of the GATA-1s isoform was also assessed by qPCR using Taqman primers as described in section 2.10.5. Reassuringly these qPCR results broadly mirror those seen with semi-quantitative PCR. In contrast to erythroid differentiation there is evidence of modulation of GATA-1s levels early in megakaryocytic differentiation with significant falls in GATA-1s expression in qPCR experiments at 6 hours after exposure to TPA in Meg-01 and K562, and at Day 2 in K562 cells (3 biological replicates per group with each sample assayed in triplicate, experiments repeated on 2 occasions). This could represent a need to down-regulate GATA-1s to allow induction of terminal megakaryocytic differentiation. However an alternative explanation is that application of TPA produces non-specific toxic effects on mRNA synthesis or stability in the cells,

manifesting as a non-specific reduction in all transcript levels. This is not supported by concurrent analysis of a housekeeping gene 18SrRNA which shows extremely stable expression levels even after TPA treatment (results shown in chapter 5.21). GATA-1FL expression also falls steeply in K562 cells although a low level rise is seen in Meg-01 (fig 3.9a). Again the different levels of progenitor cell commitment in these two cell lines may explain the different behaviours, although it is also possible that the presence of trisomy 21 in Meg-01 cells in some way prevents GATA-1FL down-regulation. These possibilities are investigated further in chapter 5.



**Fig 3.9 GATA-1 isoform expression during in vitro megakaryocytic differentiation of cell lines**

(a) Shows results of semiquantitative PCR using densitometry as described for fig 3.8 (n=1). (b) Taqman qPCR for the GATA-1s isoform shows statistically significant fall in GATA-1s expression at 6 hours and Day 2 for K562 cells and at 6 hours only for Meg-01 (n=3). \*= $p \leq 0.05$ , \*\*= $p < 0.01$

*In summary, there is evidence of dynamic regulation of GATA-1 isoforms during in vitro erythroid and megakaryocytic differentiation. GATA-1FL is down-regulated during*

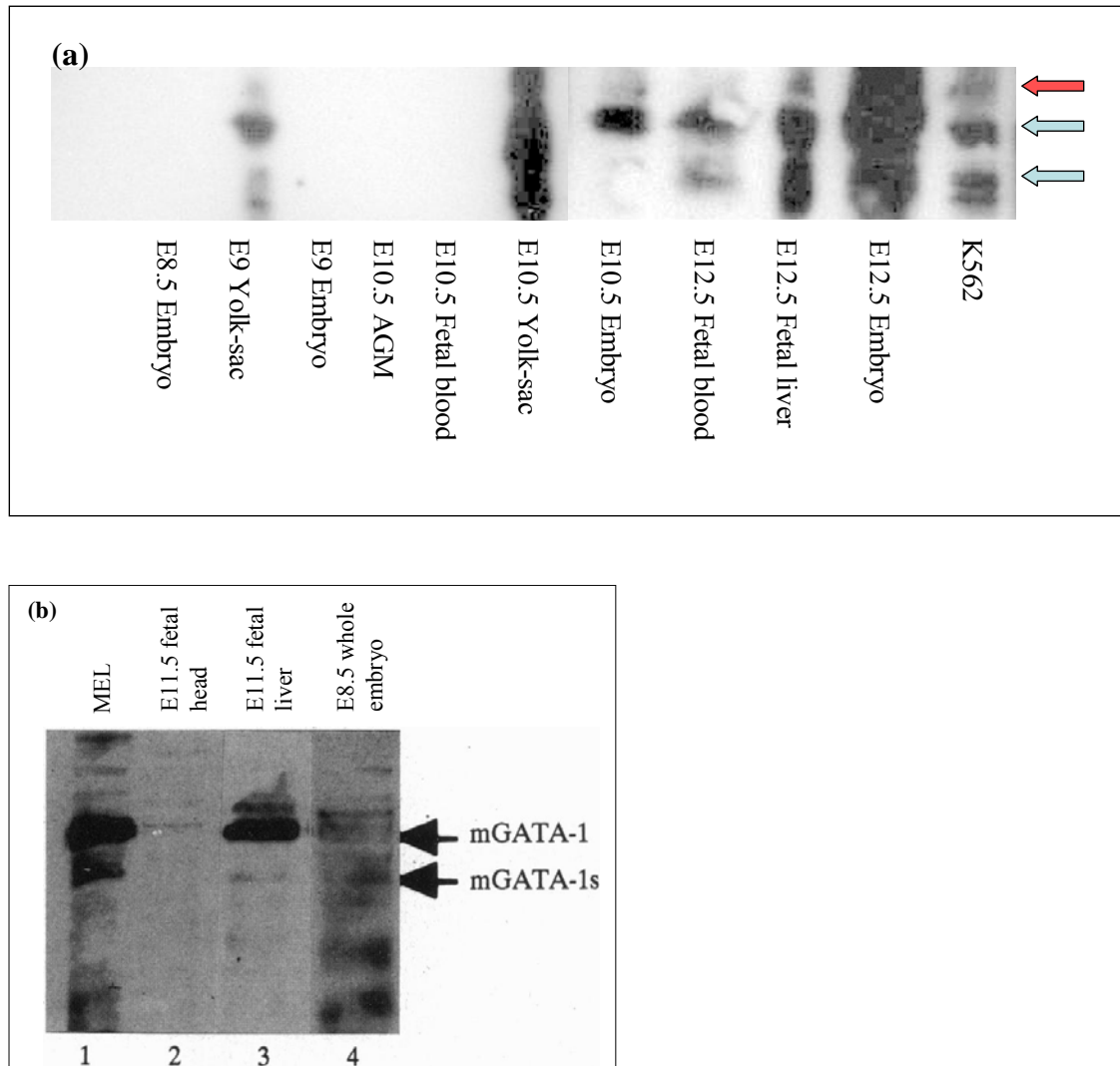
*erythroid differentiation of K562 cells, GATA-1s levels are low and do not appear to alter significantly. However, GATA-1s shows statistically significant down-regulation during megakaryocytic differentiation in both K562 and Meg-01 cell lines. These results support a role for GATA-1s in maintenance of megakaryocytic precursor populations with a need to down-regulate expression to allow terminal differentiation.*

### **3.2.5 GATA-1 isoform expression during murine embryogenesis**

In parallel with the above human studies the expression of GATA-1 isoforms was also assessed in mice. As established in section 3.1, this analysis relies on western blotting to detect the two protein products produced by alternative translation of a single mRNA.

As discussed in 1.2.8, GATA-1s has been reported to be expressed early in murine development prior to GATA-1FL expression (Calligaris, *et al* 1995). This assertion is based on a single western blot which has a lot of background making interpretation difficult. This blot is reproduced in fig 3.10b, the day 8.5 embryo sample, reported as showing exclusive GATA-1s expression, is in lane 4. No subsequent studies have addressed patterns of isoform expression early in development. In order to investigate this further, mouse embryos were obtained at E8.5, E9, E10.5, and E12.5 (gestation was confirmed by somite counting). Microdissection techniques were used to separate out the AGM region, yolk-sac, placenta and rest of body and fetal blood was also collected (see section 2.6.2 for details). These samples were red cell lysed (section 2.6.1) and then lysed in mammalian tissue lysis buffer (section 2.9.1), whole tissue extracts were used for western blotting. Results are shown in fig 3.10a. These blots are difficult to interpret due to high background, unfortunately attempts to improve clarity by the use of nuclear extracts failed. Attempts at optimising the nuclear extract protocol using mouse bone marrow were unsuccessful (data not shown) - the nuclear extract protocol in use appears to be very sensitive to the presence of contaminating red cells and is only really suitable for cell lines. This was confirmed by others (Isla Hamlett, IMM, Oxford – personal communication). A K562 whole cell extract was run as a positive control (far right-hand lane). Tentative interpretation of the blots suggests that GATA-1FL expression is seen in the E9 yolk-sac, E10.5 yolk-sac and embryo, and in all tissues at E12.5 i.e. fetal blood, fetal liver and rest of body. A number of the lanes - E8.5 embryo and E9 embryo, E10.5 AGM and Fetal blood - do not contain adequate protein (using higher band marked with a red arrow as an internal loading control) making them un-interpretable. Looking at GATA-1s expression this appears to be present in E9 and E10.5 yolk sac, the E10.5 embryo shows absent or weak expression and then GATA-1s is detectable in all embryonic tissues at E12.5. These

findings confirm early GATA-1FL and GATA-1s expression in the yolk-sac but do not support the contention that GATA-1s expression precedes that of GATA-1FL. Although the expression of GATA-1s in the yolk-sac and not embryo at E10.5 is intriguing it may merely represent differences in protein loading in these two samples, in addition the embryo at day 10.5 had already had the AGM region removed, making it a relatively poor source of haematopoietic cells.



**Fig 3.10 GATA-1 isoform expression during mouse embryonic development**

(a) Blue arrows show position of GATA-1FL and short isoforms, red arrow represents non-specific band used as loading control (b) western blot analysing GATA-1 isoform expression in early murine embryogenesis – reproduced from Calligaris et al 1995.

*Therefore, like humans, mice express both isoforms of GATA-1 early in development. These results do not confirm the published report of early exclusive GATA-1s expression but do confirm both GATA-1FL and GATA-1s isoform expression at the protein level in early embryonic tissues.*

### 3.2.6 Isoform expression in murine cell lines

Again in parallel with human studies, GATA-1 isoform expression in murine cell lines was also investigated. A number of haematopoietic cell lines were cultured in vitro with harvesting for RNA and protein (nuclear extracts). The cell lines and their putative cell of origin are shown in Table 3.2. Initially PCR was performed to look for GATA-1 mRNA – Fig 3.11a. Bearing in mind that this PCR cannot distinguish the two GATA-1 isoforms in mice, GATA-1 expression was detected at the mRNA level in cells representing haematopoietic precursors (FDCP-mix, B6SutA and 32D), erythroid precursors (MEL) and in a mast cell line (P815) but not in myelomonocytic and macrophage cell lines (Raw264, P388 and WeHi3B). It was also not seen in the murine embryonic fibroblastic cell line NIH3T3.

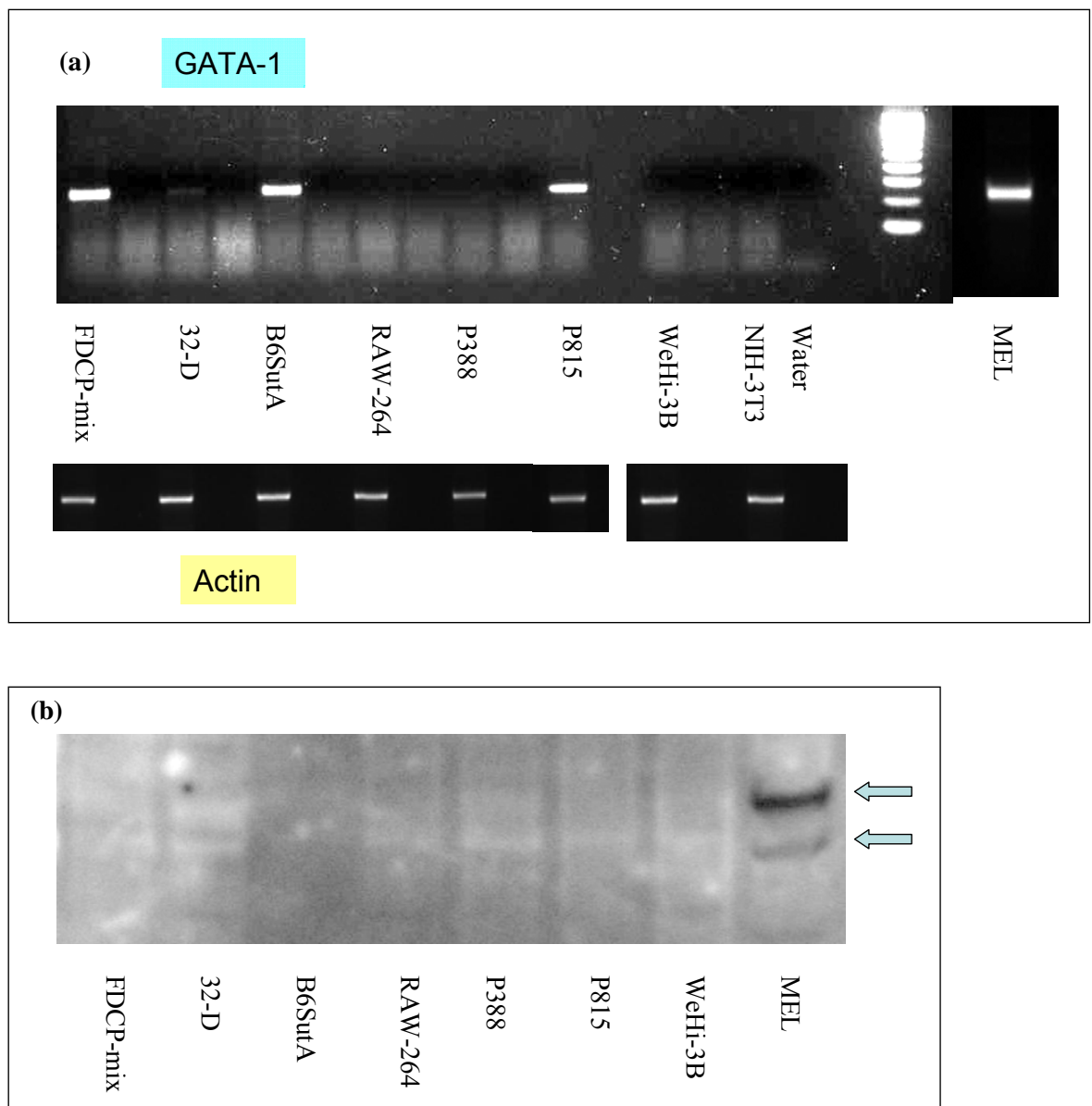
Cell line	Putative cell of origin
Factor-dependent cell mix – Patterson (FDCP-mix A4)	Multipotent progenitor
32D	Multipotent progenitor
B6SutA	Multipotent progenitor
Raw264	Monocyte/Macrophage
P388	Lymphoid macrophage
P815	Mast cell
WeHi3B	Myelomonocytic
NIH3T3	Embryonic Fibroblasts
MEL	Erythroid progenitors

**Table 3.2 Murine cell lines**

Interestingly, at the protein level GATA-1 expression was only detected in the erythroid cell line MEL which expressed both isoforms (Fig 3.11b). The bone marrow precursor lines B6SutA and FDCP mix did not show evidence of expression. This may reflect the fact that haematopoietic precursors express transcripts for transcription factors at the mRNA level but often do not produce functional protein until the cell has undergone lineage commitment – so called lineage priming (Hu, *et al* 1997) (this paper used FDCP mix as one of its model systems). The mast cell line P815 also failed to show expression at the protein level. Since mast cells are known to require GATA-1 for terminal



differentiation, rather than proliferation in culture, this again may represent post-transcriptional control of mRNA expression.



**Fig 3.11 GATA-1 isoform expression in murine haemopoietic cell lines**

(a) PCR using exon 4 and 5 primers gives a GATA-1 band of 279bp (top gel image), an actin positive control confirms adequate cDNA in all samples (b) western blot using M20 antibody, arrows show the position of GATA-1FL and GATA-1s protein bands.

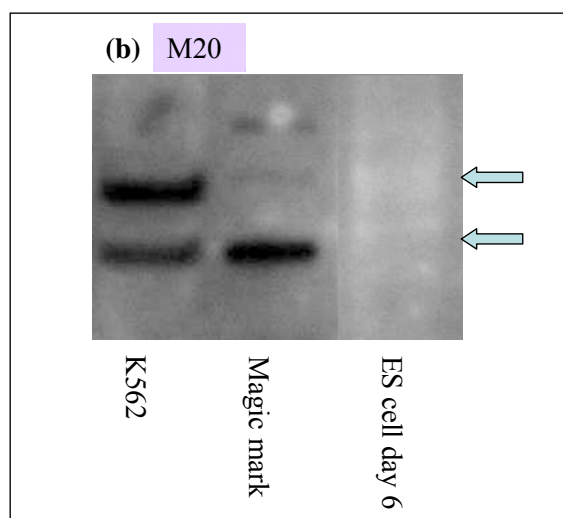
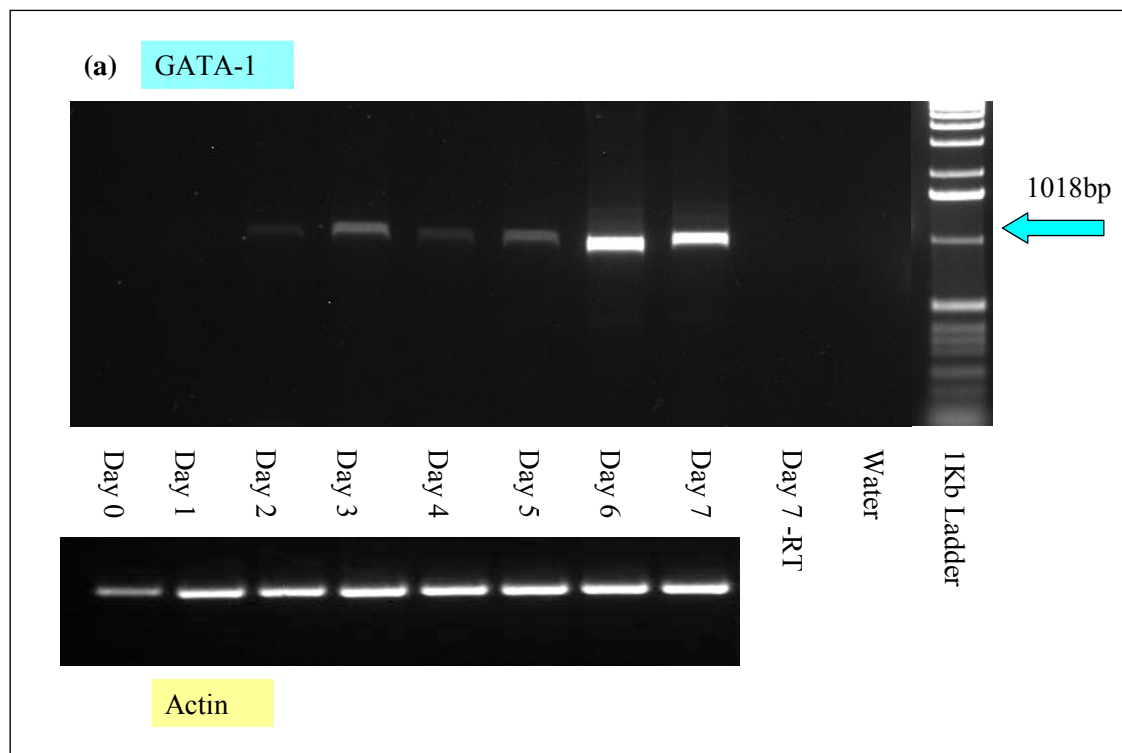
*These studies confirm GATA-1mRNA expression in haematopoietic progenitors, erythroid progenitors and mast cell populations but not in monocytic or macrophage cell lines. Detectable protein expression is only seen in erythroid precursors. This may be because of post-transcriptional regulation of GATA-1 mRNA in earlier precursor cells, or simply reflect levels of GATA-1 below the limits of detection with these antibodies.*

### **3.2.7 *GATA-1 expression during embryonic stem cell in vitro haematopoietic differentiation***

To address dynamic regulation of GATA-1 expression in murine haematopoiesis, embryonic stem cells were induced to differentiate in vitro. The use of embryonic stem cells to model haematopoietic differentiation is discussed in chapter 4. Briefly, embryonic stem cells can be maintained in a totipotent state in culture by addition of cytokines and serum, the most important component of which is leukaemia inhibitory factor (LIF). Withdrawal of LIF leads to loss of their multipotency and commitment to a variety of lineages with a particular propensity to form haematopoietic cells. Differentiation is encouraged by allowing the cells to aggregate and therefore mimicking cellular interactions during early embryogenesis. This is achieved by initial culture of cells as “hanging drops” to allow their association to form embryoid bodies followed by culture in non-adherent petri dishes in the absence of LIF. Cells can then be harvested daily and expression of genes associated with commitment to haematopoietic or other cell fates can be assayed. It has been shown that the sequence of gene expression in this system mimics that seen in vivo during haematopoietic differentiation with sequential switching on of primitive and definitive haematopoiesis.

Following formation of embryoid bodies and withdrawal of LIF (as described in materials and methods section 2.14.2.1) ES cells were harvested daily. Transcriptional profiling of GATA-1 expression was performed by RT-PCR. Fig 3.12a shows expression of GATA-1 mRNA (expected product size 1030bp), weakly detectable at day 2, increasing in intensity at day 6 and still detectable at day 7. It should be noted that there appears to be some modest variation in product sizes between the individual lanes, this was investigated and found to be due to the temporary use of SYBR Green (Invitrogen) in place of Ethidium Bromide as a DNA staining agent in the electrophoresis gel. Binding of SYBR Green to DNA leads to an alteration in molecular weight which can affect gel migration patterns (in a non-linear fashion). For this reason SYBR Green is best used to post-stain electrophoresis gels after DNA migration. All subsequent gel electrophoresis was performed with ethidium bromide. Western blotting was performed on ES cell embryoid bodies from day 6 when GATA-1mRNA expression was strongest (fig 3.12b). These failed to show expression of GATA-1 at the protein level and therefore did not allow dissection of the timing of onset of isoform expression during ES cell in vitro haematopoietic differentiation. This finding is not entirely unexpected. As mentioned above there is good evidence that during early stages of cellular commitment key transcription factors are

expressed at the mRNA level but that protein expression can be absent or at a very low level.



**Fig 3.12 GATA-1 expression during in vitro haematopoietic differentiation of murine embryonic stem cells**

(a) PCR using exon 1 (alt splice 1 forward) and exon 5 (GATA-1 reverse) primers (b) Western blot using M20 antibody. Middle lane shows magic mark size marker dark bottom band (lower arrow) = 40kDa. Blue arrows mark position of GATA-1FL and GATA-1s respectively.

### 3.3 Identification of additional GATA-1 transcripts

The experiments described above document GATA-1FL and GATA-1s isoform expression patterns in human and murine haematopoiesis. In order to comprehensively address the roles of GATA-1 isoforms in haematopoiesis it is important to establish whether additional GATA-1 isoforms exist. A number of haematopoietic transcription factors such as Erg, Runx-1 and C/EBP $\alpha$  have multiple isoforms, with evidence for differential usage of isoforms during haematopoietic ontogeny (Bedi, *et al* 2008, Corsetti and Calabi 1997, Fujita, *et al* 2001). To look for additional isoforms three approaches were taken:

1. *in silico* analysis of GATA-1 genomic sequence and EST database to look for possible additional exons and alternatively spliced transcripts.
2. Intron spanning PCR of GATA-1 cDNA at various points during ontogeny and in different lineages to identify alternatively spliced transcripts.
3. Western blotting to look for different sized proteins detected by polyclonal GATA-1 antibodies.

The results of these studies are described below.

#### 3.3.1 *In silico* analysis of GATA-1

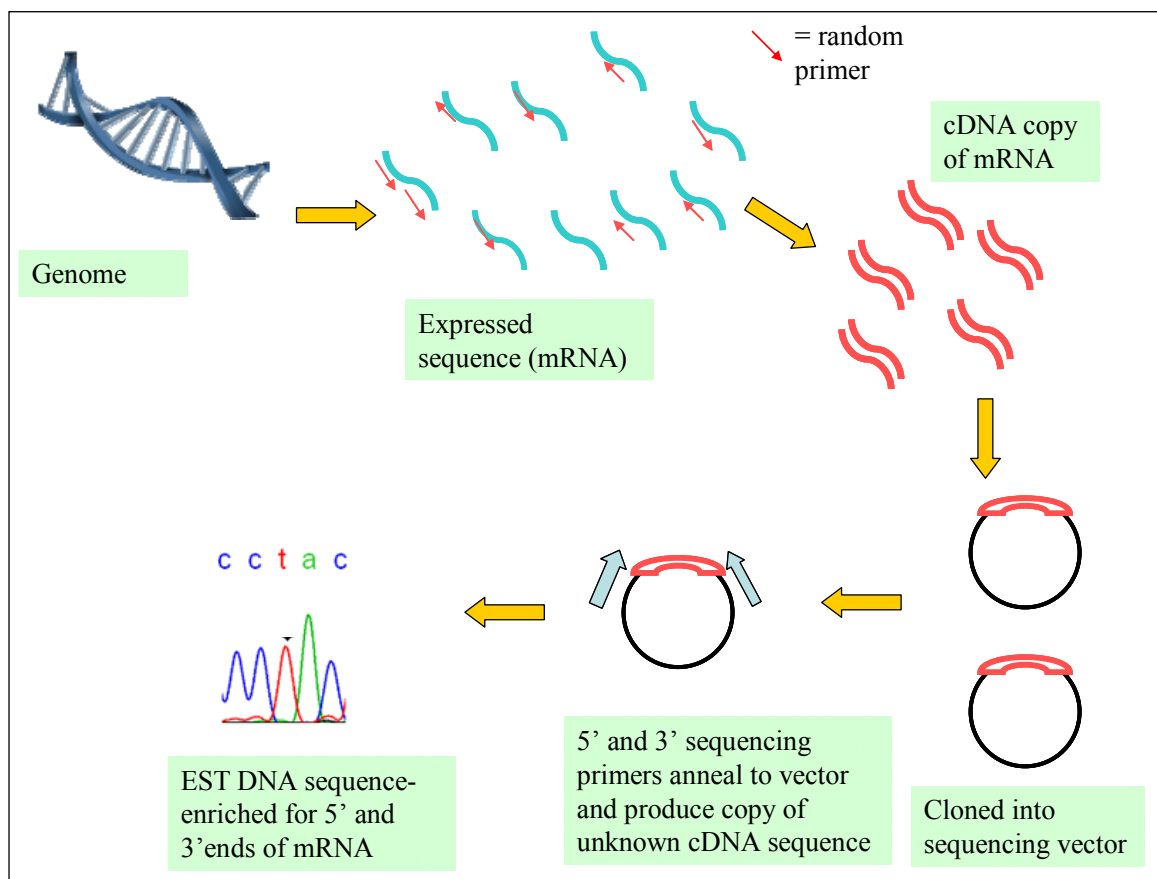
##### 3.3.1.1 GATA-1 exon analysis

A number of software programmes exist for prediction of the position of introns and exons in genomic sequences. One of the most accurate for human/vertebrate sequences is GENSCAN (<http://genes.mit.edu/GENSCAN.html>). This programme claims to identify exons with a sensitivity of 78% and a specificity of 81%, this compares to 44% and 45% for the commonly used GeneID programme (<http://genome.imim.es/software/geneid>). Using the human GATA-1 genomic sequence derived from an X-chromosome contig assembly (geneID 2623 NCBI Entrez gene: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>), Genescan accurately predicted all the known GATA-1 exons but did not identify any further exons either internally or up or downstream of the known coding sequence. The same was true for the mouse GATA-1 genomic sequence (GeneID 14460). This suggests that if additional isoforms do exist, then they will utilise different combinations of known GATA-1 exons, rather than novel exons.

### 3.3.1.2 GATA-1 EST analysis

Expressed sequence tags (ESTs) are small pieces of DNA sequence generated by sequencing of cloned cDNA derived from cellular mRNA (fig 3.13). This technique allows the identification of coding portions of the genome, discovery of new genes and discovery of alternatively spliced transcripts of existing genes (Nagaraj, *et al* 2007). Of note EST sequences tend to be short ie 200-500bp long and these sequences are not checked for sequencing errors (the whole point of EST generation is to provide a rapid, inexpensive tagging system for expressed transcripts in various tissues rather than comprehensive assessment of individual genes) so results should be interpreted with caution.

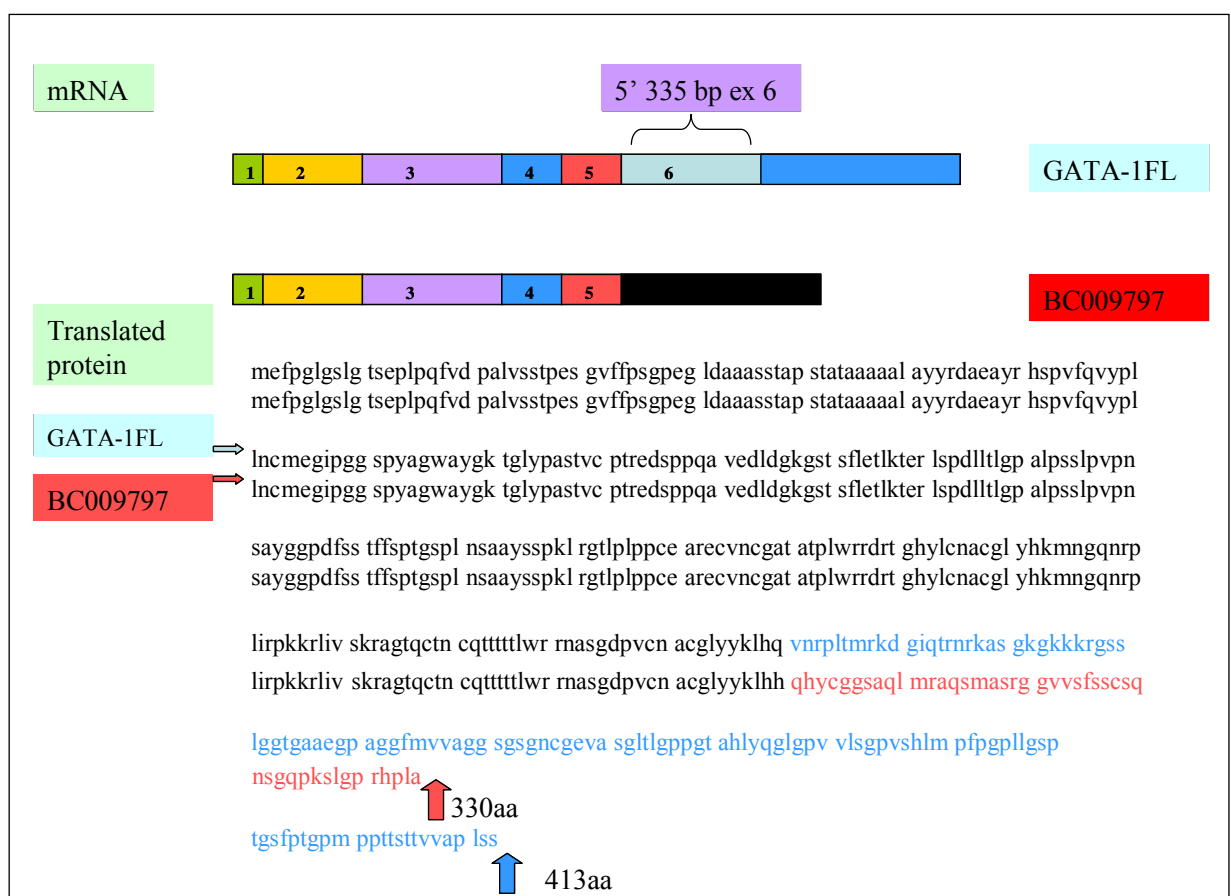
As mentioned above the GATA-1 EST database confirms the expression of an exon1/3 spliced transcript (GATA-1s) in humans but not in mice, no further alternative splicing variants were detected by this method.



**Fig 3.13 Generation of expressed sequence tags**

### 3.3.1.3 GATA-1 related sequences held on the NCBI gene database

A second approach was taken by analysing reference and related mRNA sequences for the entire GATA-1 open reading frame held in the NCBI database. Besides the verified human mRNA transcript - NM\_002049, a second transcript was listed - BC009797, which is said to encode GATA-1 mRNA. This transcript was generated by the mammalian gene collection programme (MGC) – a multi-institutional effort to identify and sequence a complete open reading frame for all human and mouse genes (Strausberg, *et al* 2002). BC009797 was sequenced from a cDNA library derived from chronic myeloid leukaemia bone marrow cells. Comparing BC009797 against NM\_002049 reveals that this transcript lacks the first 335bp of exon 6 and the new exon 5/6 splice site results in a frameshift producing a shortened protein with a completely different predicted C-terminal amino acid sequence. This is illustrated in fig 3.14. Following this in silico analysis the reference clone was obtained for further in vitro investigation as detailed in section 3.4.



**Fig 3.14 Comparison between NM\_002049 and BC009797**

3.3.2 RT-PCR analysis of mRNA

Two PCR strategies were used to detect additional GATA-1 transcripts and these are shown in fig 3.15. Primers designed to anneal to exon 1 and 6 should produce bands corresponding to GATA-1FL and GATA-1s in humans. Any additional splicing variants, as long as they incorporated exon 1 and 6 would produce additional bands of different length. Fig 3.16a shows GATA-1FL and GATA-1s bands but no additional sized products in human K562 cells. Since GATA-1 is already known to utilise a different first exon in Sertoli cells it is possible that use of an exon 1 forward primer may miss some isoforms lacking this exon, therefore primers were also designed to anneal to exons 3 and 6 (fig 3.15). Use of these primers in samples from patients with CML, normal controls and the cell line K562, confirmed the presence of the expected 1094bp exon3/4/5/6 product but also produced a large band of approximately 1500bp (fig 3.16b). To establish its identity, this larger band was excised from the gel, TOPO cloned into pCR2.1 (see materials and methods, section 2.12) and commercially sequenced ([www.agowa.de](http://www.agowa.de)). Sequencing revealed that it was composed of exons 3-5 plus intron 5 and exon 6 (fig 3.16c). It therefore is likely to represent a splicing intermediate rather than a functional alternative isoform.

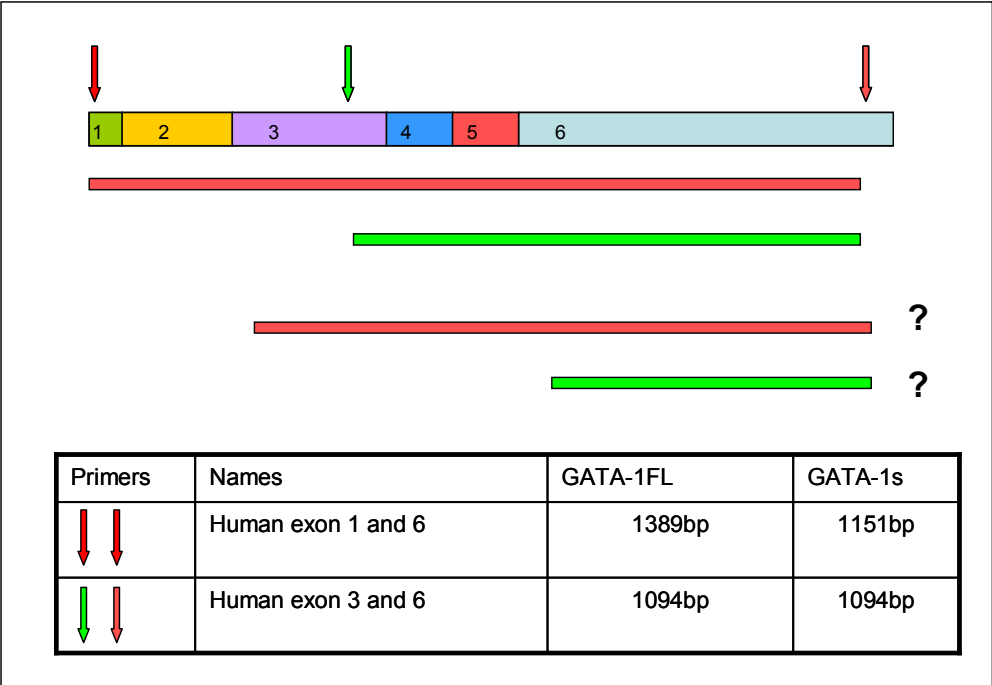
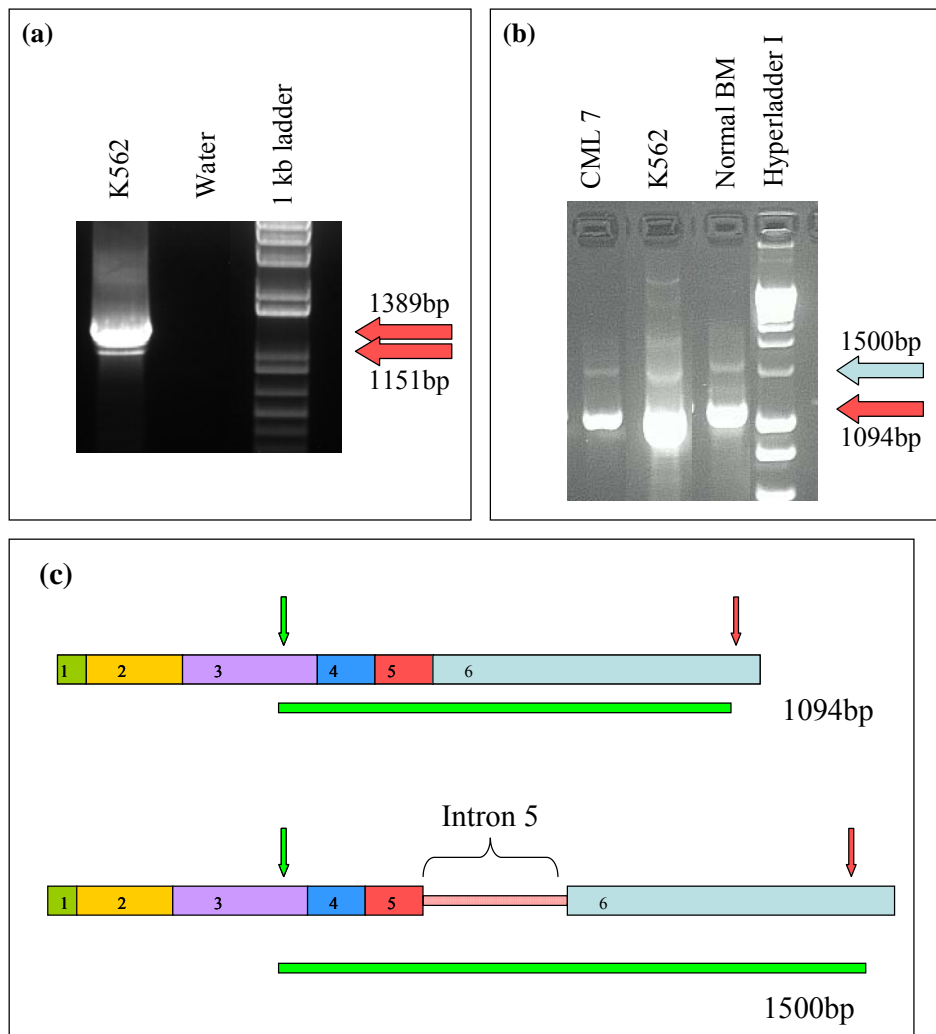


Fig 3.15 PCR strategies to detect novel GATA-1 transcripts

Annealing sites of exon 1, 3 and 6 primers shown with expected product sizes of GATA-1FL transcript (red bar) and GATA-1s transcript (green bar), question marks indicate possible additional products if alternative isoforms exist.



**Fig 3.16 Detection and analysis of an additional 1500bp transcript in human samples**

### 3.3.3 Western blotting reveals only two bands

Alternative isoforms can also be separated on the basis of molecular weight of their corresponding protein. As discussed in 3.1.2 use of the polyclonal M-20 GATA-1 antibody detects two isoforms by western blot corresponding to the full-length and short isoform, whilst the N-terminus N-6 antibody identifies a single band (fig 3.3b). M-20 is raised against the C-terminus of GATA-1FL, since GATA-1 proteins have highly conserved zinc fingers in the C-terminus it is unlikely that any functional isoform would lack this domain and therefore escape detection by this antibody. Even postulating that an isoform might lack DNA binding activity, conservation of the N-terminus (and therefore detection by the N6 antibody) would have to occur for the protein to bear any resemblance to GATA-1. Therefore it would be predicted that any alternative isoforms should be detected by a two-pronged approach targeting antibodies to the N- and C-termini. Western blotting of a



variety of tissues in both mouse and humans as shown throughout this chapter (figs 3.3b, 3.4, 3.6, 3.7, 3.10, 3.11, 3.12 and 3.19) failed to reveal additional bands attributable to additional isoforms.

*In summary, in silico, PCR and western blotting approaches were used to try and identify any additional GATA-1 isoforms. Two additional GATA-1 transcripts were identified. Firstly, a cDNA clone with an altered C-terminus – BC009797 isolated from CML bone marrow. Secondly, an elongated exon 3-6 PCR product corresponding to an exon3/4/5-intron5-exon6 splicing intermediate. The latter is unlikely to be functionally significant but the former warranted further investigation as discussed below.*

### **3.4 In vitro analysis of BC009797**

The MGC BC009797 clone could represent: (i) an additional physiological isoform of GATA-1 (ii) a mutant transcript (with either widespread or sporadic occurrence in CML) or (iii) an artefact. In order to distinguish these possibilities, further investigations were performed. All MGC clones are available for public distribution via the IMAGE consortium (Integrated Molecular Analysis of Genomes and their Expression) whose primary goal is to create arrayed cDNA libraries and associated bioinformatics tools, and make them publicly available to the research community. Therefore in order to assess the possible relevance of the BC009797 transcript the corresponding IMAGE clone was ordered (MGC:13628, IMAGE:4048082) (Geneservice, Cambridge UK) and analysed.

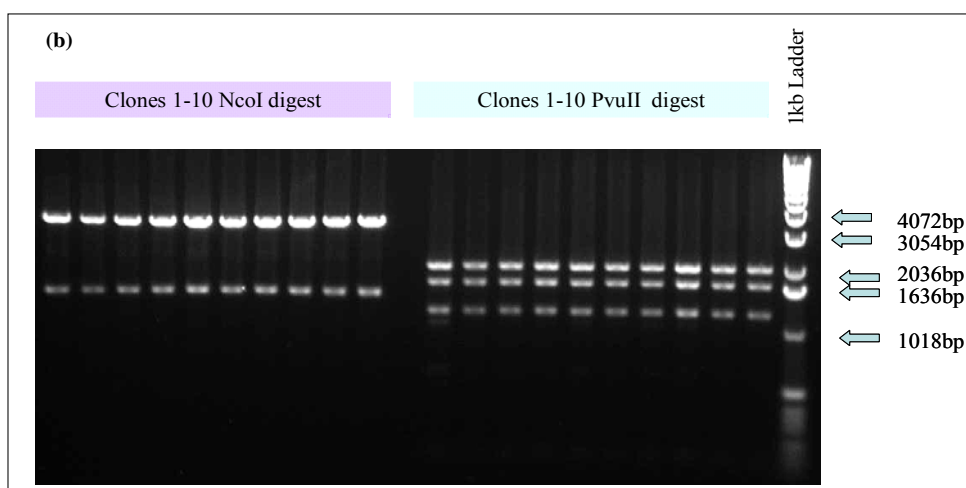
#### **3.4.1 Analysis of IMAGE clone reassigns cloning vector but confirms sequence**

The IMAGE clone was supplied as a streak on LB Agar with chloramphenicol. This clone was re-streaked onto an LB Agar plate containing 27µg/ml chloramphenicol and incubated at 30°C for 24 hours to produce single colonies. 10 single colonies were picked, grown overnight in LB broth with chloramphenicol and DNA was extracted using a commercial miniprep technique (Qiagen – see materials and methods, section 2.12.2.3). IMAGE clones are supplied in a number of different cloning vectors depending on the laboratory that initially isolated them. The datasheet indicated that BC009797 was cloned into the pOTB7 cloning vector. A predicted restriction enzyme site map was constructed using this data and the restriction enzymes NcoI and PvuII were chosen to analyse BC009797 orientation within this vector and confirm the identity of the supplied clone. As seen in fig 3.17 the

restriction enzyme fragments produced from this diagnostic digest do not correspond to the predicted sizes in either orientation (fig 3.17a & b). A single DNA prep was sent for sequencing using the T7 sequencing primer whose sequence is present in the majority of cloning vectors. The results of this sequencing confirmed the presence of BC009797 cDNA, ruling out the possibility that the wrong clone had been despatched. Analysis of the short stretch of associated vector sequence revealed similarity with another commonly used MGC cloning vector – pDNRLIB. Construction of a theoretical RE site map of BC009797 in this vector led to prediction of fragment sizes (fig 3.17a) that exactly matched those seen with the NcoI PvuII digest (fig 3.17b). Therefore the identity of the cloning vector was reassigned from pOTB7 to pDNALIB. Fortunately, both these vectors carry chloramphenicol resistance genes.

(a)

Vector	RE	Correct orientation	Wrong orientation
pOTB7	NcoI	2574	1590
		419	1403
	PvuII	2003	1638
		990	1355
pDNRLib	NcoI	3672	2668
		1460	2444
	PvuII	2049	1684
		1695	2060
		1244	1244
		144	144



**Fig 3.17 Restriction enzyme analysis of IMAGE clone suggests an alternative cloning vector.**

(a) predicted sizes of digest products (b) diagnostic digests were performed as described in section 2.7.3. Blue arrows indicate size markers in 1kb ladder (Invitrogen). NcoI digest produces fragment sizes of approximately 3500bp and 1500bp, PvuII digest produces fragments of approximately 2000bp, 1700bp and 1200bp. These fragment sizes equate to predicted values for pDNRLib-BC009797 digest in correct orientation as shown in (a)

Extended sequencing of 1700bp from the T7 priming site confirmed that the full BC009797 sequence exactly matched the NCBI database sequence and therefore that the skipping of the 5' end of exon 6 and resultant frameshift were not due to sequencing artefacts

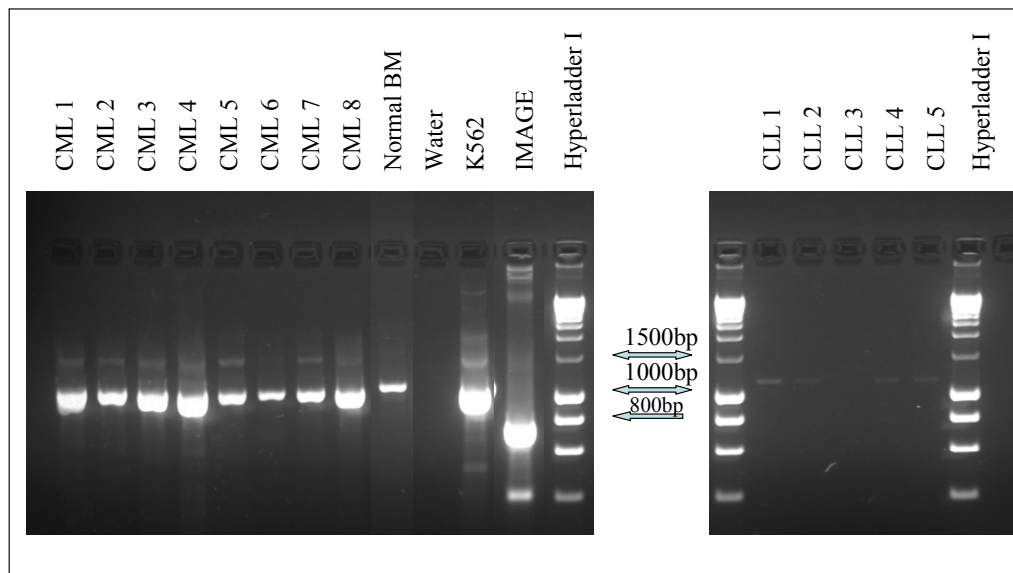
*In summary, these results confirm that the published BC009797 sequence is correct, ruling out the possibility that BC009797 C-terminal alterations were a sequencing artefact.*

### **3.4.2 Analysis of BC009797 expression at the mRNA level**

Having established that BC009797 appears to be a true transcript, further investigations set out to look at whether this transcript was widely expressed and whether it had functional consequences.

Since BC009797 was originally cloned from the bone marrow of a patient with chronic myeloid leukaemia (CML) it was decided to use primary CML samples to look for expression of this transcript in other individuals. cDNA generated from bone marrow mononuclear cell mRNA samples from 9 patients with CML at diagnosis or relapse, and for comparison, 5 patients with chronic lymphocytic leukaemia (CLL) and one normal bone marrow sample, was obtained (kind gift Dr Paz Carreno and Dr Andrew Clarke, Glasgow University). PCR was performed with primers annealing in exon 3 and the 3' end of exon 6. As shown in Fig 3.15, these primers would be predicted to give a 1094bp band with the GATA-1FL and GATA-1s isoforms but produce a shorter 765bp band in the presence of a BC009797 transcript. The IMAGE expression plasmid was used as a positive control and confirmed the ability of the primers to amplify a 765bp fragment. No other samples amplified this product but instead all produced the 1094bp fragment (fig 3.18). The larger 1500bp band seen on these gels corresponds to the intermediate splicing variant discussed in section 3.3.2 and also shown in Fig 3.16. The existence of detectable GATA-1 transcripts, albeit at low level, in CLL samples is surprising and is presumed to reflect a low level of red cell or myeloid cell contamination in these samples.

In addition, an in silico analysis of the human EST database using a query term that encompassed the novel exon 5/6 splice site junction was performed. Despite being only 150bp from the 3' end of the mRNA this analysis did not pick up any cloned sequences corresponding to this variant.



**Fig 3.18 PCR screening for BC009797-like transcript in other human primary cells**

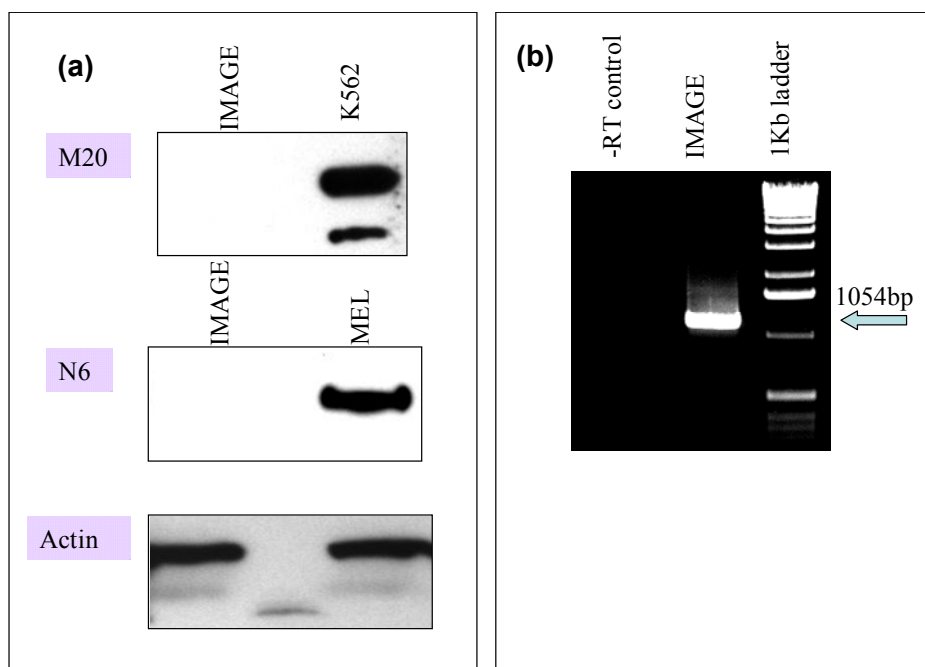
*Overall, these results suggest that the BC009797 variant is not widely expressed in CML. However, sample sizes were too small to establish whether this transcript occurs at low frequency in this condition, or whether it is likely to represent a unique patient specific mutation.*

### **3.4.3 A functional analysis of BC009797**

Analysis of the functional consequences of the C-terminal changes in BC009797 was performed to establish whether this transcript was likely to be of pathological or physiological significance. In order to analyse BC009797 at the protein level it was recloned into a vector known to produce good expression levels in haematopoietic cell lines – pEF6/V5HisTOPO (Invitrogen)- as discussed in chapter 5.2.5.1. This was achieved by design of forward and reverse primers corresponding to the 5' and 3' ends of the gene – ensuring all coding sequences including the stop codon were included – (sequences listed in section 2.4.2) and PCR using ReddyMix followed by gel purification and TOPO cloning (see materials and methods, section 2.10 and 2.12). After overnight growth on LB Agar with Ampicillin (50µg/ml), 6 colonies were picked and subsequent miniprep and restriction digest confirmed the presence of the BC009797 insert in the correct orientation in 5/6 clones. One of these was maxi-prepped (section 2.12.2.3) and subsequent sequencing of this clone (pEF6IMAGE) confirmed that the PCR amplification had not introduced any errors.

### 3.4.3.1 Western blotting is unable to detect BC009797

pEF6IMAGE, was transfected into HEK293 ( $6 \times 10^5$  cells seeded on a 6-well plate) using a commercial transfection reagent – Polyfect (Qiagen). Cells were harvested at 48 hours and a nuclear extract was made (section 2.9.2). Western blotting using M20 (polyclonal, C-terminus) and N-6 (N-terminus) revealed no band in the extract from IMAGE transfected cells, despite bands on the positive control (K562 for M20, MEL for N6), fig 3.19a. This is consistent with the lack of N6 binding to human GATA-1 (as discussed in 3.1.2.2) and suggests that M20 may be unable to bind to the aberrant C-terminus predicted for BC009797. However the possibility that lack of detection was due to lack of expression from the vector cannot be ruled out by this experiment. To look for production of mRNA transcripts by the vector RT-PCR was performed using GATA-1 exon 1 and 6 primers (sequence listed in 2.4.1), these primers would be predicted to give a 1054bp product with the IMAGE transcript. Fig 3.19b confirms the presence of a 1054bp transcript. The absence of a band in the –RT control excludes the possibility that this band is due to contamination with vector DNA. These findings are supported by reporter gene analysis, discussed in section 3.4.3.2 below, which shows that at least partially functional GATA-1 was being made by the vector (fig 3.20).

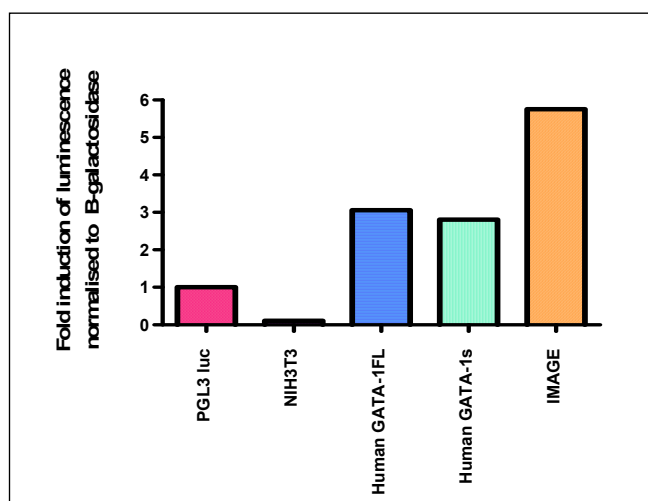


**Fig 3.19 BC009797 mRNA and protein expression**

(a) Western blotting with M20 and N6 GATA-1 antibodies reveal positive control bands (K562 and MEL respectively) but no band with the IMAGE expression vector transfected into HEK293 cells. Adequate protein loading is confirmed by the actin control (b) PCR using human GATA exon 1 and 6 primers should produce a band of 1054bp as shown. The minus (-) RT control confirms this was not due to contaminating vector DNA.

### 3.4.3.2 Reporter assays suggest BC009797 has enhanced transactivation potential

The lack of BC009797 protein detection by commercially available antibodies could imply that the mRNA was not transcribed into protein or that the altered C-terminus abolished binding of the M20 antibody. An alternative approach, using a GATA-1 reporter assay to detect GATA-1 protein production, was taken. This approach has the advantage of not only detecting protein production by the vector but also assessing its functional ability. The reporter plasmid pGL3hGATA-1 (kind gift from Dr P Vyas, WIMM, Oxford) contains cis-acting GATA-1 DNA binding elements. Binding of functional GATA-1 protein at these sites results in activation of transcription, leading to the expression of firefly luciferase. Sections 2.13.1 and 5.2.4.3 describe this reporter assay methodology in more detail. pEF6IMAGE was cotransfected into NIH3T3 cells alongside pGL3hGATA-1 and a  $\beta$ -galactosidase producing vector to normalise for transfection efficiency. Optimisation of this technique is discussed in chapter 5.2.4.3. After 48-72 hours cells were harvested, lysed and the supernatant analysed for  $\beta$ -galactosidase activity (Promega) and luciferase activity (Promega). Fig 3.20 shows that pEF6IMAGE was able to produce a 5-6 fold induction of reporter activity over background. This was higher than the induction of expression from GATA-1FL and GATA-1s human plasmids which showed 2-3 fold induction over background.



**Fig 3.20 Function of BC009797 in a luciferase reporter assay**

Graph shows fold induction of luminescence in NIH3T3 cells co-transfected with the test plasmid (Human GATA-1FL, Human GATA-1s or IMAGE), a luciferase reporter plasmid and a  $\beta$ -galactosidase producing plasmid. Results were normalised to  $\beta$ -galactosidase activity to control for transfection efficiency and expressed as fold induction compared to reporter plasmid alone (lane 1 PGL3 luc). Background levels of luminescence from untransfected NIH3T3 cells were negligible (lane 2 NIH3T3). Results from one representative experiment out of three shown.

The fact that GATA-1FL and GATA-1s both only produce low level induction of reporter activity over background is surprising given that previous reports (albeit using different reporter constructs) suggest that GATA-1FL should produce considerable enhancement of reporter activity (up to 5 -10 fold) compared to GATA-1s (Wechsler, *et al* 2002). The design and testing of these human GATA-1 plasmids is discussed further in chapter 5. However, the ability of BC009797 to induce luciferase expression both confirms that it encodes a functional protein and raises the intriguing possibility that it might have enhanced DNA binding activity. BC009797 diverges from GATA-1FL and GATA-1s at position 959 and the GATA-1 C-terminal zinc finger sequence ends at position 970 so the majority of the C-Zinc finger remains intact. The function of the C-terminus beyond the zinc finger is largely unknown although Runx-1 has been shown to bind in this region (Elagib, *et al* 2003). It is possible that the distal C-terminus region interacts with transcriptional repressors, or is the site of post-translational regulation of protein activity, and that the altered BC009797 terminus escapes from these influences. Further investigation of this poorly defined region is warranted.

***Overall, these results suggest that the BC009797 transcript encodes a functional protein. This protein is undetectable by commercially available antibodies, but has demonstrable GATA-1 like activity on reporter assays. These assays also suggest it may actually have enhanced trans-activation ability. This implies that it may have had a functional consequence for the original patient from whom it was cloned. It also highlights the possibility that the distal C-terminus may be an important, previously underappreciated, functional domain of GATA-1.***

#### **3.4.4 Summary of the relevance of BC009797**

A number of observations made above suggest that BC009797 is unlikely to be a physiological isoform of GATA-1 i.e.

1. The expression of this transcript has not been detected in cell lines or patient samples (other than the founder patient) by PCR (fig 3.18), or by in silico analysis of EST databases.
2. The existence of an alternative shorter exon 6 is not predicted by in silico Genescan analysis.
3. It is very unusual for alternatively spliced isoforms to produce frameshifts.

Despite this two important findings emerged from these studies:

1. The existence of this transcript in a patient with chronic myeloid leukaemia along with its possible increased activity suggests that BC009797 may represent a mutated aberrantly functioning form of GATA-1 rather than a normal isoform. As discussed above, this mutation may have had functional consequences for the patient. These observations led to further investigation of possible GATA-1 mutations in CML – the results of which are presented in Chapter 6.
2. Investigation of this transcript has shown a possible enhanced transactivation potential associated with an altered distal C-terminus. The function of this region of GATA-1 requires further investigation.

### **3.5 Chapter 3 summary**

Overall this chapter describes experiments designed to assess (i) the mode of GATA-1 isoform expression in humans and mice (ii) temporal and spatial expression patterns of GATA-1 isoforms during haematopoietic ontogeny (iii) expression patterns in different lineages (iv) whether additional GATA-1 isoforms are detectable. With the aim of uncovering patterns of expression that may shed light on the role of GATA-1s and/or novel isoforms in haematopoietic development. The following conclusions can be drawn:

1. There is good evidence for alternative splicing as a method of isoform production in humans but no evidence of this mechanism in mice. In contrast mice appear to produce GATA-1s by alternative translation of a single mRNA by a mechanism known as “leaky scanning”.
2. GATA-1s expression seems to accompany GATA-1FL expression in the majority of tissues both during ontogeny and in different cell lines. On only one occasion was GATA-1s production seen to be more prominent than GATA-1FL, this was in a human primary monocyte population, this result was not duplicated in an independent sample or in monocytic cell lines although GATA-1s expression could be detected in some of these alongside GATA-1FL. GATA-1s expression in monocytes, eosinophils and mast cells has never previously been reported.



3. GATA-1s expression does not appear to be dynamically regulated during erythroid differentiation but rapid down-regulation of expression occurs on induction of megakaryocytic differentiation.

4. Two novel GATA-1 transcripts were identified in these studies. One represents a likely intermediate splicing form of doubtful significance. The other - BC009797- appears to be a mutated form of GATA-1 with an altered C-terminus. Intriguingly this transcript showed increased activity in reporter gene assays, suggesting that the distal C-terminus may be an important functional domain of GATA-1. Its occurrence in CML raises the possibility that this disease may be associated with aberrant GATA-1 expression – a hypothesis tested in chapter 6.

Further investigation of the differences between GATA-1FL and GATA-1s at a functional level will be described in the following chapters.

## 4 GATA-1 gene targeting in murine embryonic stem cells

---

### 4.1 Aims

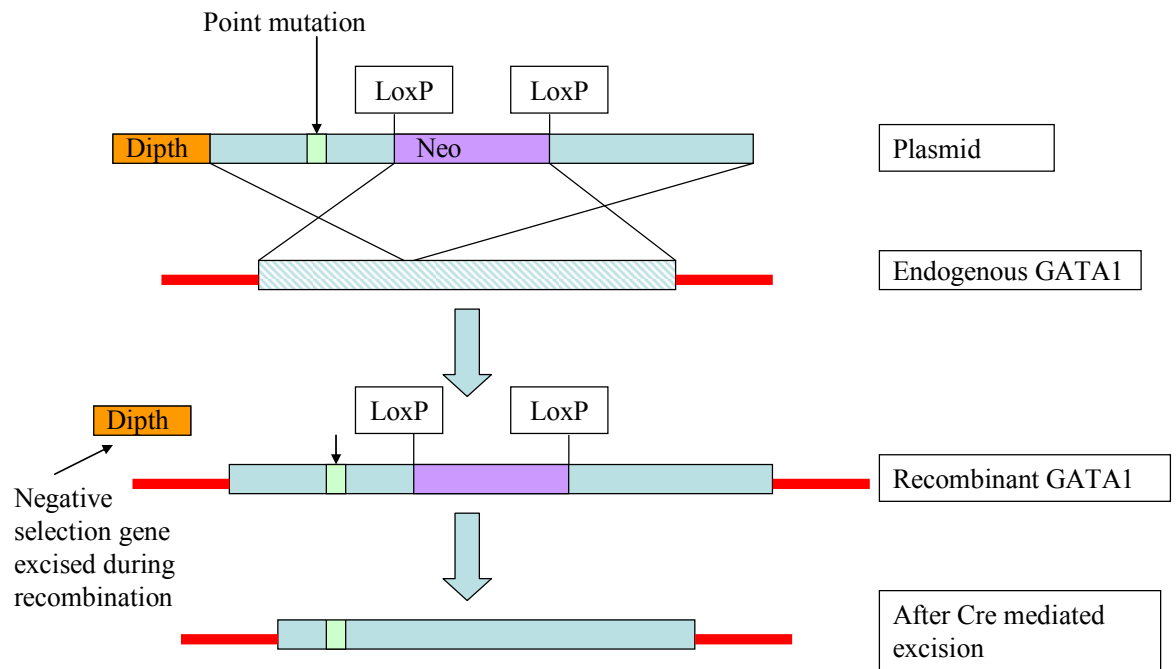
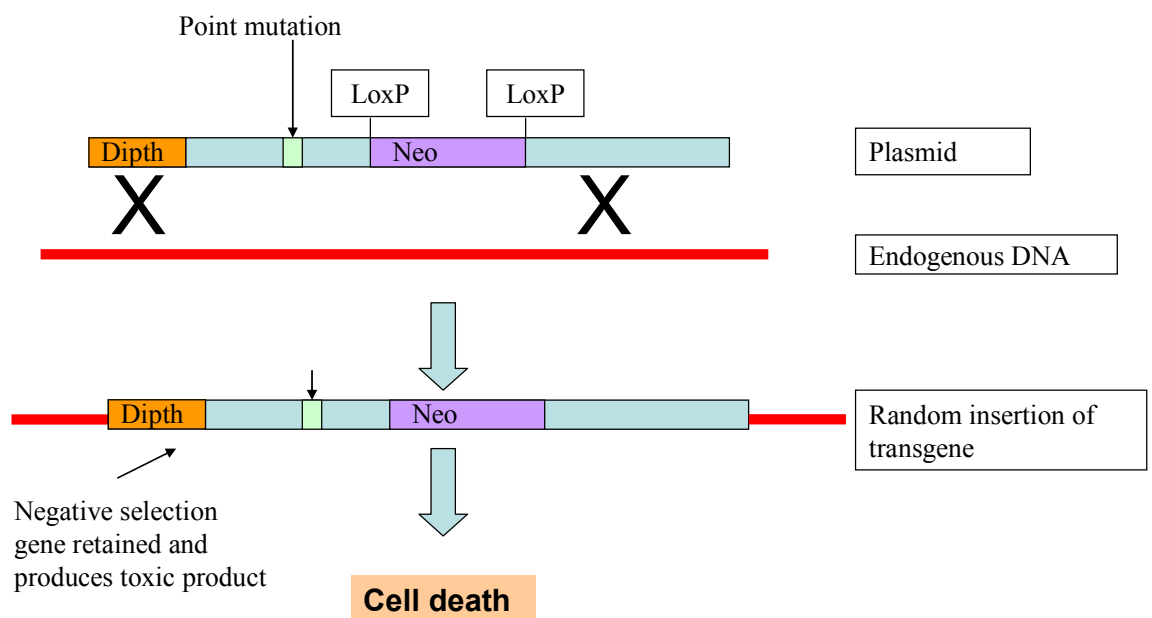
In order to assess any functional differences between the two GATA-1 isoforms it was decided to look at their ability to drive haematopoietic differentiation *in vitro*. This chapter describes attempts to engineer murine embryonic stem (ES) cells to exclusively express either GATA-1FL or GATA-1s isoforms. These cells could then be used to study the ability of the two isoforms to independently support primitive and definitive haematopoietic colony formation using *in vitro* haematopoietic differentiation in methylcellulose (Keller, *et al* 1993). In addition, transcriptional profiling could be used to look at isoform specific regulation of gene expression (Baird, *et al* 2001). Although these targeted ES cells could, in theory, be used to make transgenic mice (Bradley, *et al* 1984), this aim was beyond the time scale and funding parameters of this project.

### 4.2 Introduction

Assessment of gene function is greatly aided by transgenic technology, allowing introduction of cloned copies of a gene of interest into a cell and evaluation of the impact of expression of this transgene on cellular processes. These transgenes can be over-expressed from an exogenous promoter or can target the endogenous gene either inactivating it, or replacing it with an altered version. Gene over-expression studies, described in chapter 5, are very useful tools to study gene function. These approaches provide information on the cellular pathways that the target gene interacts with and the molecular consequences of gene expression. However, use of an exogenous promoter produces unregulated expression. Because transcription factors show significant concentration dependent effects, and are tightly temporally and spatially controlled, these studies must be interpreted with caution. An artificial promoter/enhancer construct for GATA-1, termed GATA-1 HRE has been devised that encompasses all the regulatory elements required to express a downstream gene in the same temporal and spatial pattern as GATA-1, as shown by recapitulation of GATA-1 expression patterns using a LacZ reporter gene (McDevitt, *et al* 1997a). Although use of this construct overcomes some of the problems associated with over-expression vectors, the cell still possesses its own endogenous copy of GATA-1. This means that it is impossible to look exclusively at the

function of a single isoform, as the cell still retains the capacity to express the other isoform from the endogenous locus. Furthermore, expression of the additional GATA-1 gene from the transgenic construct will alter the stoichiometry of transcription factors in the cell.

A more physiological way to study the biological effects of GATA-1 isoform expression would be to target the endogenous GATA-1 locus and replace the normal GATA-1 gene with a construct capable of exclusive GATA-1FL or GATA-1s expression – a process known as gene targeting (Capecchi 2001). This approach is sometimes termed ‘knock-in’ technology i.e. the endogenous gene is replaced by an alternative gene (as opposed to ‘knock-out’ where gene expression is ablated from the endogenous locus). Although in this particular case it could equally be thought of as an isoform specific ‘knock-out’. Replacement of the endogenous gene is achieved by exploiting the biological process of homologous recombination (Alberts 2008). This involves genetic exchange between two similar or identical strands of DNA. Homologous recombination is the method by which cells repair double stranded breaks in DNA, and also the mechanism by which sister chromatids cross-over at meiosis allowing random assortment of genetic information. Introduction of exogenous DNA with large regions of homology with the endogenous locus can induce the cell to exchange the exogenous and endogenous DNA, albeit at low frequency. Although technically difficult, use of this approach has been aided by a number of advances in molecular biology, most importantly the ability to culture embryonic stem cells (ES cells) in vitro and maintain them in an undifferentiated totipotent state. Since over 10 million cells can be grown in a single 10cm culture dish even extremely low frequency events (such as homologous recombination) will occur within this cell population. The difficulty comes in being able to identify these events. This is aided by positive selection strategies to narrow down cells that have incorporated the targeting construct and negative selection strategies to enrich for cells that have undergone homologous recombination (Capecchi 2001). This must be followed by a sensitive screening strategy to identify the cells that have undergone the desired recombination event. Once targeted cells have been identified it is possible to excise the positive selection marker by utilising site specific recombinases (such as Cre-recombinase) that recognise a target sequence (LoxP sites) flanking the selectable marker. The principles behind this method of gene targeting are illustrated in Fig 4.1 and described in more detail below.

**(a) Homologous recombination****(b) Random Insertion****Fig 4.1 Outline of gene targeting approach**

(a) Introduction of a point mutation by using a positive (neomycin resistance gene- neo) and negative (diphtheria toxin) selection strategy, followed by Cre recombinase mediated removal of the selectable marker. (b) If the targeting construct randomly integrates, expression of the negative selection marker will result in cell death.

## **4.2.1 Targeting selection strategies**

### **4.2.1.1 Positive selection**

Introduction of an antibiotic resistance gene into the middle of the targeting construct, as long as it is flanked by significant lengths of homologous DNA, will not prevent homologous recombination occurring (Hasty, *et al* 1991). Cells that have integrated foreign DNA can then be selected for by virtue of their resistance to antibiotics.

The commonest antibiotic resistance gene is neomycin phosphotransferase (*neo*) which confers resistance to the aminoglycoside antibiotic G418 by promoting G418 phosphorylation. For reliable expression of this gene in ES cells, the mouse Pgk-1 gene promoter is usually used (Nagy 2003).

### **4.2.1.2 Negative selection**

Although positive selection helps narrow down the cell population to those that have incorporated the targeting construct, the frequency of random insertion is many orders of magnitude higher than that of homologous recombination (Mansour, *et al* 1988, Thomas and Capecchi 1987). Both scenarios will confer antibiotic resistance. Initial experiments to establish methods for gene targeting chose target genes with a selectable phenotype. For example, the hypoxanthine phosphoribosyl transferase (*Hprt*) gene on the X-chromosome was used – which confers resistance to 6-thioguanine (6-TG) treatment when mutated (Hooper, *et al* 1987, Thomas and Capecchi 1987). Therefore cells which had undergone homologous recombination with a construct that inactivated the endogenous *Hprt* gene could be selected for by exposure to 6-TG. In order to extend gene targeting to non-selectable genes it is necessary to have an alternative method to enrich for homologous recombination events. One possibility is to incorporate a second selectable marker at the distal end of the targeting construct. Cells that randomly integrate the construct will express this marker since the vast majority of random insertion events occur at the ends of a linearised molecule (Roth, *et al* 1985). In contrast, cells that undergo homologous recombination will excise it as it lies out with the area of homology (see fig 4.1). If this marker confers a negative selection pressure (i.e. kills cells it is expressed in) then its use alongside a positive selection marker will allow preferential survival of cells that have undergone homologous recombination rather than random insertion. This positive/negative selection strategy was first described by Mansour *et al* (Mansour, *et al* 1988). A number of negative selection markers have been used, the most common is Herpes simplex virus

thymidine kinase which confers sensitivity to the antiviral agent Gancyclovir. Alternatively the diphtheria toxin gene, derived from *Corynebacterium diphtheriae*, can be used (Awatramani, *et al* 2001, Soriano 1999). This gene product inhibits chain elongation in protein synthesis via its inactivation of elongation factor-2 (EF-2). It is extremely toxic to cells and has the advantage of not requiring addition of further drugs to the culture medium.

## **4.2.2 Optimal design of the targeting vector**

### **4.2.2.1 Homology arms**

The length and degree of homology between the exogenous and endogenous DNA are critical for efficient recombination. There should be at least 6-8kb of homologous sequence (Hasty, *et al* 1991) and although the two arms flanking the selectable marker can be of unequal length, the minimum size for an arm is 1kb. The two sequences should be as identical as possible. Homologous recombination using a targeting construct derived from isogenic DNA is 20-fold more efficient than if the DNA is derived from an alternative mouse strain (te Riele, *et al* 1992). To ensure this, the targeting DNA should be amplified from cells from the same mouse strain as the ES cell parental DNA – most commonly this is the 129 inbred strain. In addition, for knock-in strategies, the desired genetic variation should be as close as possible to the selectable marker to minimise the chances of homologous recombination occurring between the selectable marker and the mutation which would make the locus appear targeted when in fact the desired mutation has not been introduced.

### **4.2.2.2 Removal of the selectable marker**

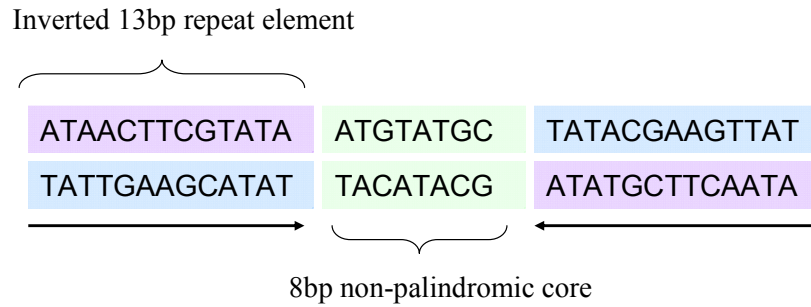
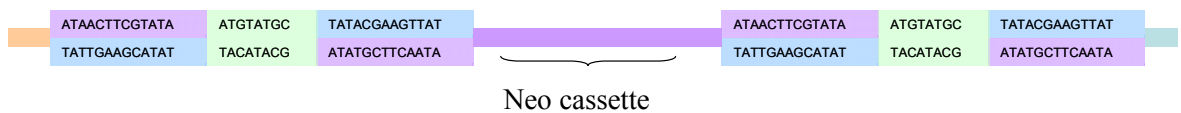
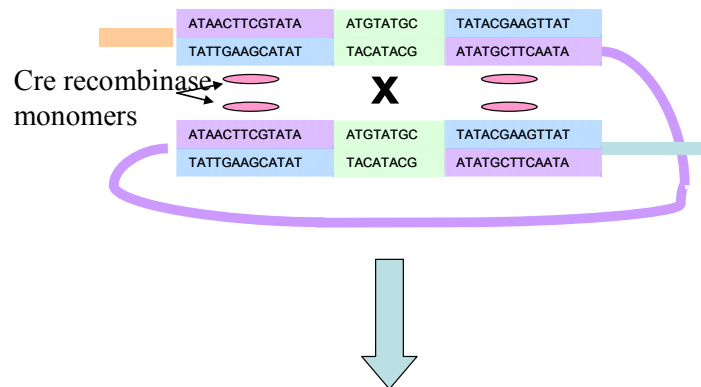
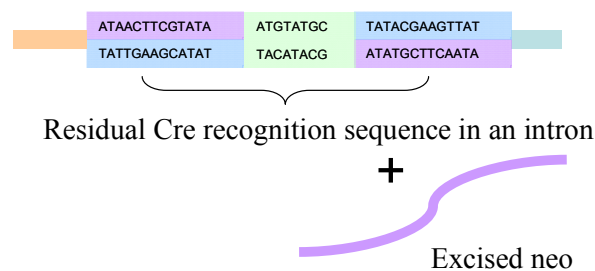
The above approaches are sufficient to produce a gene knock-out by replacing the endogenous locus with the neo antibiotic resistance gene. For knock-in approaches where the endogenous locus is to be replaced by an altered gene product more sophisticated methods are needed. Although in theory placing the selectable marker in an intron would still allow production of the knock-in gene product (i.e. the mRNA would splice out the neo gene during processing), in practise the presence of a foreign gene in an intron severely reduces mRNA production from that allele (Meyers, *et al* 1998, Nagy, *et al* 1998). In addition the neo sequence could introduce cryptic splice sites leading to truncation of the target gene. Therefore, if more subtle changes are to be introduced, it is important to be able to excise this antibiotic resistance gene once targeted clones have been identified. In

order to achieve this, the neo gene must be placed within an intron and be flanked by target sequences for site specific recombinases whose mode of action are described below:

Site specific recombinases direct recombination between two consensus DNA sequences resulting in excision of the intervening DNA. They can be utilised to allow post-insertional changes to the targeted DNA sequence. Cre recombinase is derived from bacteriophage P1 and is the most widely used recombinase for murine gene targeting. It catalyses recombination between two 34bp long *lox-P* sites which comprise two 13bp palindromic sequences flanking an 8bp core sequence as shown in fig 4.2. The latter is asymmetric and therefore confers orientation on the loxP sites. Cre protein monomers bind the consensus sequence and mediate strand cleavage and exchange. This technique has revolutionised the ability to produce subtle changes in genes such as point mutations, which cannot be achieved by knock-out techniques.

#### **4.2.3 Methods to screen for recombination events**

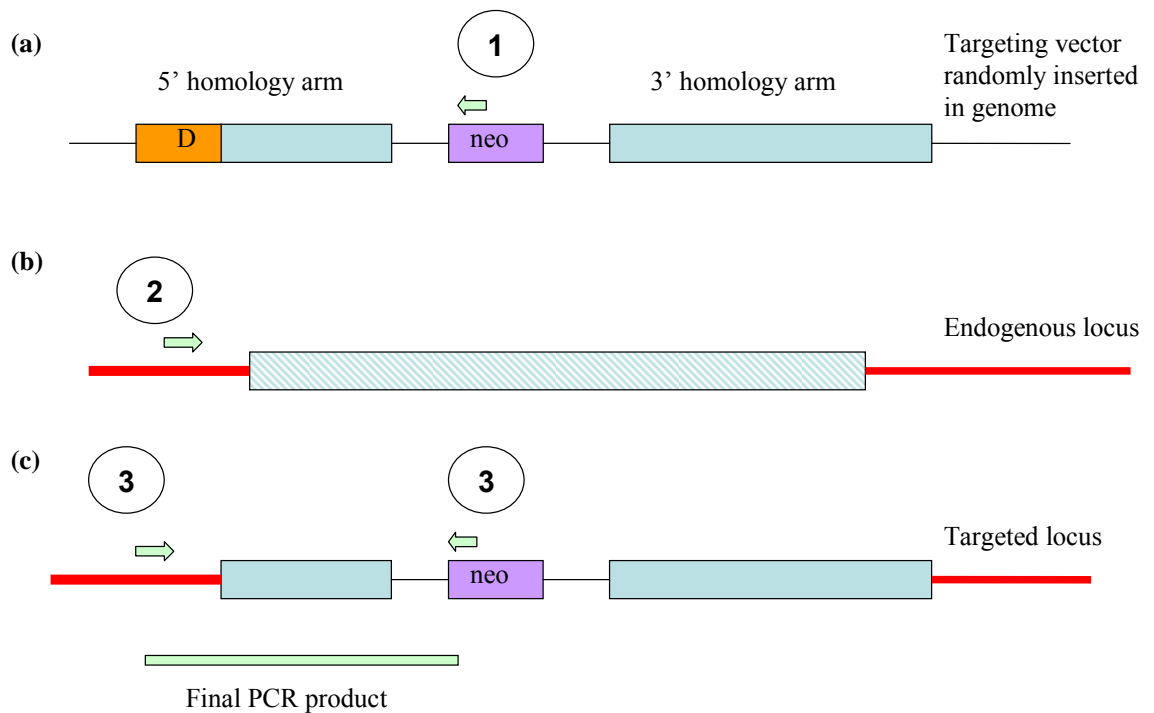
Using the two-pronged approach of positive and negative selection it is possible to narrow down the number of cells that have integrated the targeting construct and enrich for those that may have undergone homologous insertion. However, this event still remains rare – looking at published targeting frequencies for GATA-1, correct targeting is estimated to occur in the region of 1 in 100 cells remaining after positive and negative selection (Pevny, *et al* 1991). Therefore, it is necessary to devise a screening strategy capable of identifying correctly targeted cells amongst all the other cells that have undergone random insertion. The two main methods are PCR based screening and southern blotting.

**(a) Lox-P sequence elements****(b) Lox-P sites flanking the antibiotic resistance gene****(c) Cre-recombinase mediated excision****(d) After excision****Fig 4.2 Cre Lox-P mediated recombination to excise the selectable marker****4.2.3.1 PCR**

PCR screening is rapid and relatively simple but often technically difficult as the targeted gene is present at low frequency amongst a lot of relatively impure genomic DNA. It is important to devise a sufficiently sensitive technique with adequate positive controls. The method relies on designing primers to amplify a fragment unique to the targeted locus.



This is usually achieved by having one primer binding to a sequence within the antibiotic resistance gene (not present in the wild-type endogenous locus) and the second primer binding to a sequence just upstream of the targeting construct shorter homology arm (and therefore not present in the targeting vector). Therefore a product will only be amplified when the upstream sequence and the antibiotic resistance gene are brought into proximity. If the product is the correct size this situation should only arise if the endogenous locus has been targeted. This is illustrated in Fig 4.3



**Fig 4.3 Principles of PCR screening for targeted colonies**

(a) Illustrates the situation during random insertion. The 3' PCR primer, pale green arrow labelled 1, will anneal to the antibiotic resistance gene, but the 5' primer will not anneal in close proximity, resulting in no amplification. (b) At the endogenous locus the 5' primer will anneal (labelled 2) but there is no antibiotic resistance gene and therefore the 3' primer is not present. (c) In the presence of a targeted locus the 3' and 5' primers will be brought into close proximity (labelled 3) and result in an identifiable PCR product.

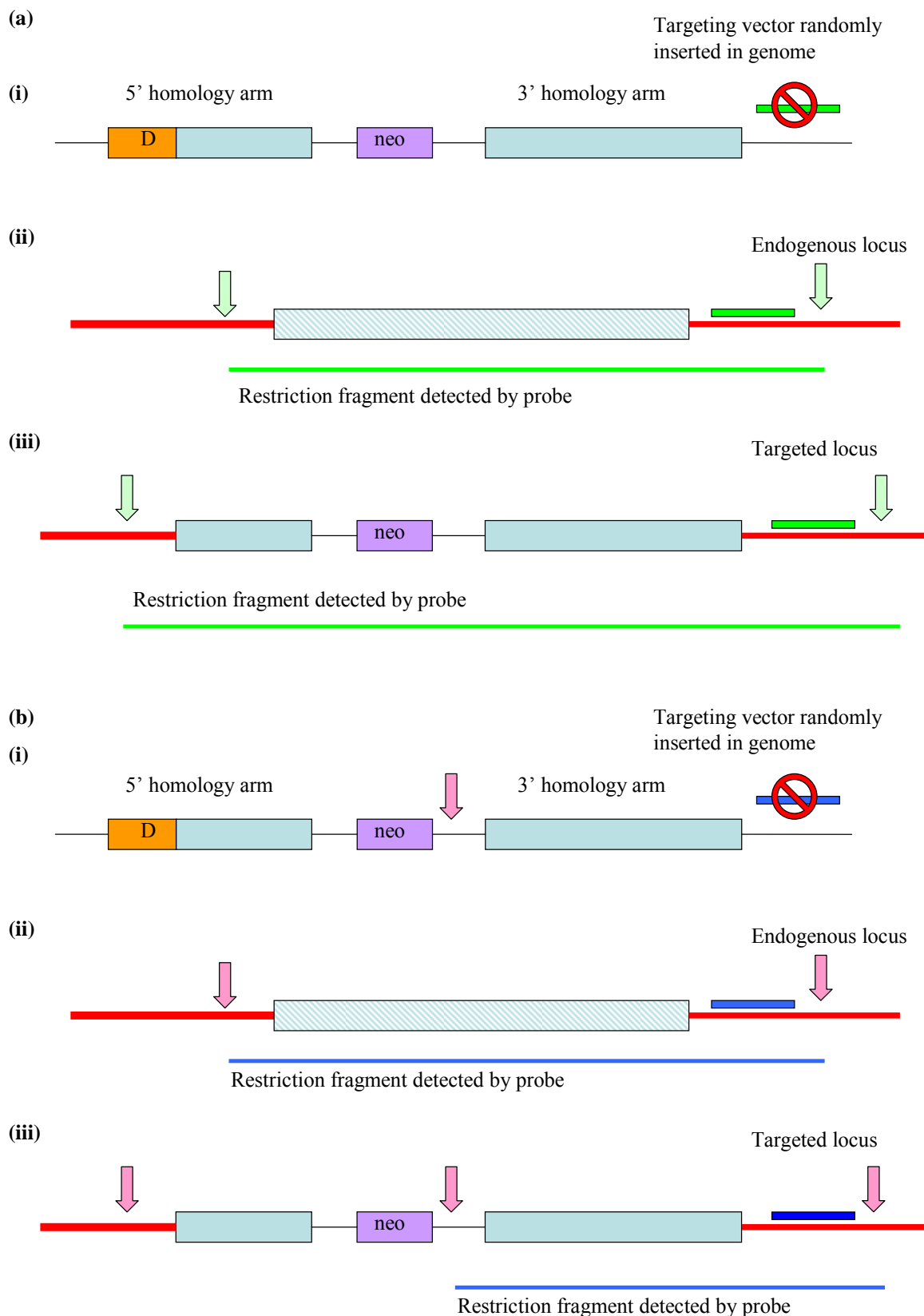
Since targeting is a rare event it is important to have a positive control to establish that lack of a targeted band is due to lack of targeting rather than technical limitations of the procedure. This is achieved by constructing an artificial plasmid containing the upstream sequences, not included in the targeting vector, and the antibiotic resistance gene in the correct position and orientation. It is important to take special care that the positive control plasmid never contaminates the PCR reaction mixture leading to false positive reactions. This can be verified by ensuring that the control vector has slight sequence variation

compared to the targeting vector and therefore any positive bands can be investigated to see if they originate from the targeted endogenous locus or the control vector. This is discussed further in section 4.8.3.1 below.

### 4.2.3.2 Southern blotting

A lot of screening strategies still employ Southern blotting as the most reliable method for identifying correct targeting events. Targeted DNA is recognised by using an external probe (i.e. annealing to the target gene but outside the region present in the targeting vector) to detect a variation in the size of restriction fragments produced by digestion of the targeted or the endogenous locus. Two possible strategies can be used, as illustrated in Fig 4.4. Firstly a restriction enzyme that cuts in the endogenous DNA either side of the targeting construct integration site will produce a larger band in the presence of a targeted allele due to the presence of the selectable marker (Fig 4.4a). This strategy has the advantage of being able to distinguish a correctly targeted locus from one that has undergone integration of the entire vector without correct homologous recombination or that has undergone integration of a concatamerised vector. However the resultant fragment is often too large to be accurately resolved by Southern blot (ideally the fragments should be 5-10kb in size), and the occurrence of suitable restriction enzymes that do not cut within the targeting vector (which has been constructed from several vectors containing artificial multiple cloning sites) is relatively rare. Alternatively, a second strategy is to identify an enzyme that cuts within the selection gene of the targeting vector (and therefore not in the corresponding endogenous locus) and also just outside one of the homology arms (Fig 4.4b). This produces a shorter fragment with the targeted locus and a larger fragment with the endogenous locus. However, this method is unable to distinguish correct homologous recombination from other events such as integration of the entire targeting vector or formation of concatamers prior to integration. To overcome this it is necessary to do an additional Southern blot using a probe at the opposite end of the targeting construct.

When selecting restriction enzymes it is important that they are not sensitive to CpG methylation as genomic DNA extracted from ES cells is usually methylated, this may restrict the choice of suitable enzymes for digestion.



**Fig 4.4 Principles of Southern blot analysis for targeted colonies**

(a) Using a restriction enzyme that cuts either side of the region targeted by the construct, results in a longer restriction fragment with the targeted locus due to the additional presence of the antibiotic resistance gene. (b) If the targeting construct introduces a new restriction site then the presence of the targeting vector will produce a smaller fragment.

## 4.3 Creation of GATA-1 isoform targeting constructs

Following the general discussion of the principles of gene targeting above, this section will go on to address the specific methods used to generate targeting vectors capable of producing exclusive GATA-1FL or GATA-1s expression from the endogenous locus. Targeting of GATA-1 is facilitated by its position on the X-chromosome. All experiments used E14 embryonic stem cells originally derived from strain 129 mouse blastocysts (Hooper, *et al* 1987). The use of these ES cells, derived from a male embryo, means that only one allele needs to be targeted to produce the desired genetic change.

### 4.3.1 Generating homology arms

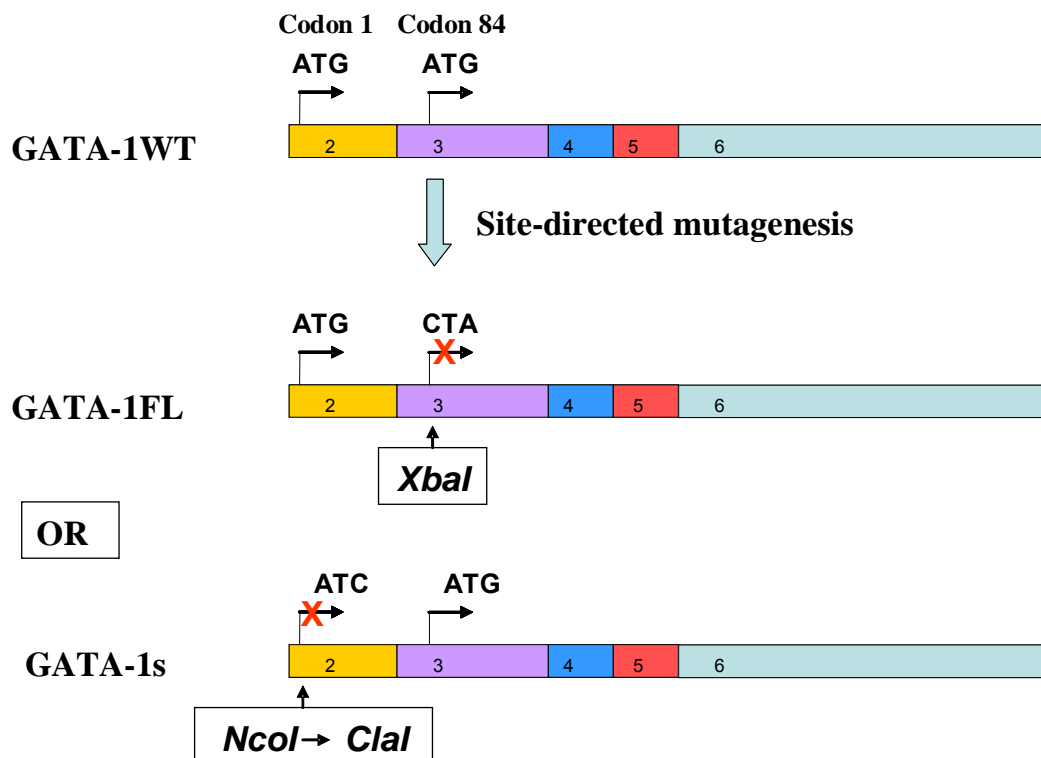
Because, as discussed above, even small variations such as point mutations can drastically reduce targeting efficiency, a high fidelity DNA polymerase with proof reading activity – Pfu-turbo (Invitrogen) was used to amplify the targeting arms. The template DNA was isogenic with the ES cell strain to be targeted (both 129). Once amplified, the products were run on a gel, the correct sized band excised, ‘A’ overhangs added, and the product was then TOPO cloned into the appropriate cloning vector. Following colony picking and minipreps, restriction enzyme (RE) digests were performed to identify clones containing the correct insert in the correct orientation. Informative restriction enzyme digests were chosen with the help of commercial DNA cloning software (Clone Manager Version 8.0, Scientific & Educational Software). The entire sequence of a single clone for each construct was obtained using primer walking methodology to produce a DNA sequence of sufficient length (Agowa.de). A problem arose when analysing the sequence of the amplified cloned product as the major gene reference bank sequence (Entrez Gene, NCBI) is derived from the C57Bl6 strain rather than the 129 strain. Fortunately the GATA-1 gene appears highly conserved between these strains as discussed below.

### 4.3.2 Introduction of point mutations into first or second start codons

Targeting vectors capable of exclusive GATA-1FL or GATA-1s production were designed by performing site directed mutagenesis of either the first or second ATG start codons therefore preventing translation of the GATA-1FL or GATA-1s protein respectively. Experiments described in the next chapter (chapter 5 section 5.2) used reporter gene assays to show that the GATA-1FL gene product carrying a mutated ATG at codon 84 displays

equivalent activity to wild type GATA-1 in its ability to transactivate luciferase expression. Since the mutation required to produce GATA-1s occurs upstream of its coding sequence it would not be predicted to introduce any functional differences in to the final protein product.

Once the homology arms described above had been amplified up and their sequence confirmed, they then underwent site directed mutagenesis of the first or second ATG to produce the two isoforms as illustrated in Fig 4.5. The exact base pair change to ablate the second ATG was carefully chosen to produce an amino-acid change with similar biochemical properties (hydrophobic, neutral) to minimise any possible effects on protein folding or function. In addition, base-pair changes were chosen since they also produced alterations in restriction enzyme sites. Therefore screening for colonies containing the desired mutation in the targeting construct was straightforward.



**Fig 4.5 Production of isoform specific targeting vectors**

For GATA-1FL a methionine to leucine (ATG to CTA) change, and for GATA-1s a methionine to isoleucine (ATG to ATC) change, were chosen. This introduced an XbaI site for GATA-1FL, whilst for GATA-1s an NcoI site was abolished and a ClaI site created.

### **4.3.3 Addition of selectable markers**

Having cloned, sequenced and performed site directed mutagenesis on the two homology arms, selectable markers were added to the targeting vector. The positive marker was the Neo gene (driven by the mouse Pgk-1 promoter), flanked by Lox-P sites to allow subsequent excision by Cre-recombinase. The Diphtheria toxin gene also driven by the Pgk-1 promoter was used for negative selection (Awatramani, *et al* 2001, Soriano 1999).

### **4.3.4 Devising and testing a screening strategy for targeted clones**

It is very important to devise a robust screening strategy for correctly targeted colonies *before* finalising the target construct design as screening can be challenging and require modifications to the targeting vector such as preferential use of certain restriction enzyme sites in the multiple cloning sequence to ensure informative digests can be produced. In addition it is important to test the screening strategy using positive controls before undertaking large scale screening of ES cell clones. Screening strategies are shown for the two targeting approaches in the relevant sections below and illustrated in Fig 4.7. Both PCR and Southern blot screens were devised. In order to test PCR primers and to provide a positive control a PCR control vector was constructed comprising the 5' arm of the targeting construct along with the flanking endogenous 5' sequence and the neo resistance gene. This vector was carefully constructed with a small variation in the neo/5' junction compared to the real targeting construct. This allowed differentiation between false positive PCR reactions due to low level contamination with control vector and the product produced by endogenous targeting. Southern blot probes were also amplified, cloned and sequenced and untargeted E14 cells were used as a positive control for the wild type band. 5' and 3' external probes were chosen. The initial screen was with a BamHI digest which cut either side of the targeting construct and within the selection marker (see Fig 4.7b). These restriction enzyme fragments were detected with a 3' external probe. An additional check was provided by an AflIII digest which cuts either side of the targeting vector and therefore produces a larger fragment with the targeted allele due to the presence of the neomycin selection cassette. These fragments were detected with a 5' probe (Fig 4.7c). Therefore this comprehensive screening strategy was able to identify potentially targeted alleles, check that both 3' and 5' recombination had occurred successfully and ensure that results were not due to integration of the whole vector or concatamer formation. Optimisation of these screening techniques is discussed in the relevant sections below.

## 4.4 Targeting construct version 1

The strategy behind the first attempt at producing a targeting vector is shown in fig 4.6. Vectors containing the positive (pgk-neo) and negative (Rosa diphtheria toxin) selectable markers were a kind gift from Dr D Gilchrist, Division of Immunology, Infection and Inflammation, Glasgow. In brief, construction of the final targeting vector comprised 5 steps:

**Step 1:** Amplification of a long 4480bp 3' homology arm comprising exons 3-6 with intervening introns. An EcoRV site was added to the 5' primer. This product was TOPO cloned into the commercial pCR4 vector (Invitrogen) (Vector A –pCR4exon 3-6). Fig 4.6a.

*Site-directed mutagenesis was performed on this construct to delete the second ATG (GATA-IFL targeting construct) discussed in 4.4.1.1.*

**Step 2:** Amplification of the short 5' arm comprising of 2344bp of exon 2 and its flanking introns with Xho1/Kpn1 restriction sites added to the 5' and 3' primers respectively. This product was also TOPO cloned into the pCR4 vector (Vector B – pCR4exon 2). Fig 4.6b.

*Site-directed mutagenesis was performed on this construct to delete the first ATG (GATA-Is targeting construct), discussed in 4.5.1.1.*

**Step 3:** Pgk-1 neo was excised from a Pgkneo vector using kpn1/notI restriction enzymes and ligated into vector B using the same restriction sites to produce vector C (pCR4exon 2/neo). Fig 4.6c.

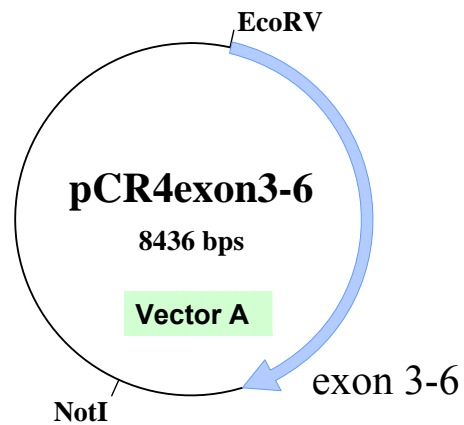
**Step 4:** The 3' arm was subcloned from vector A to vector C using EcoRV/NotI restriction sites to produce vector D (pCR4exon 2-6/neo). Fig 4.6d.

**Step 5:** The diphtheria toxin negative selectable marker was excised from a Rosa diphtheria toxin vector using XhoI/SalI and ligated into a XhoI Site in vector D to produce vector E (pCR4exon2-6/neo/diphth). Fig 4.6e.

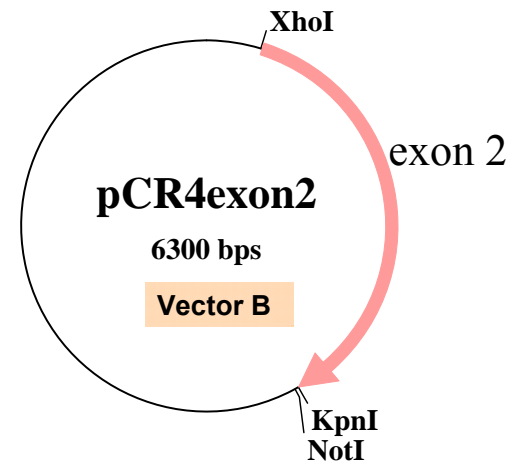
The vector could then be linearised with NotI (Fig 4.6e), prior to electroporation into E14 ES cells.

The screening strategy to identify targeted clones is shown in fig 4.7.

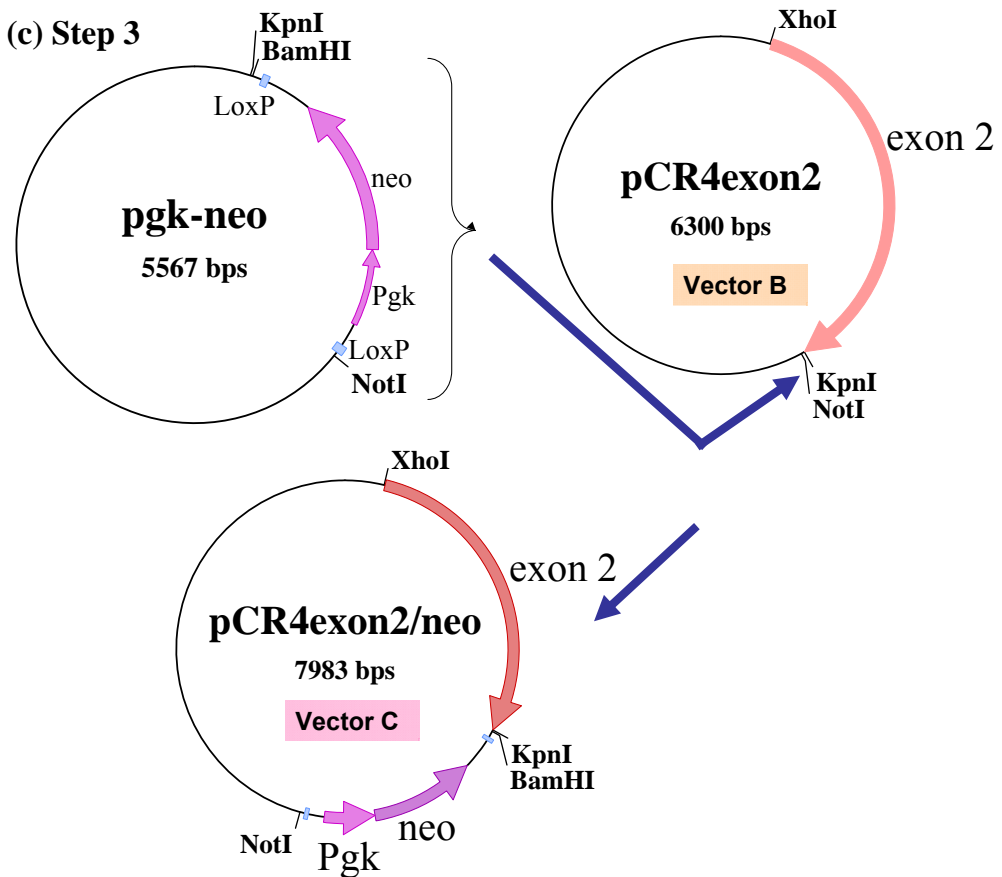
## (a) Step 1



## (b) Step 2



## (c) Step 3

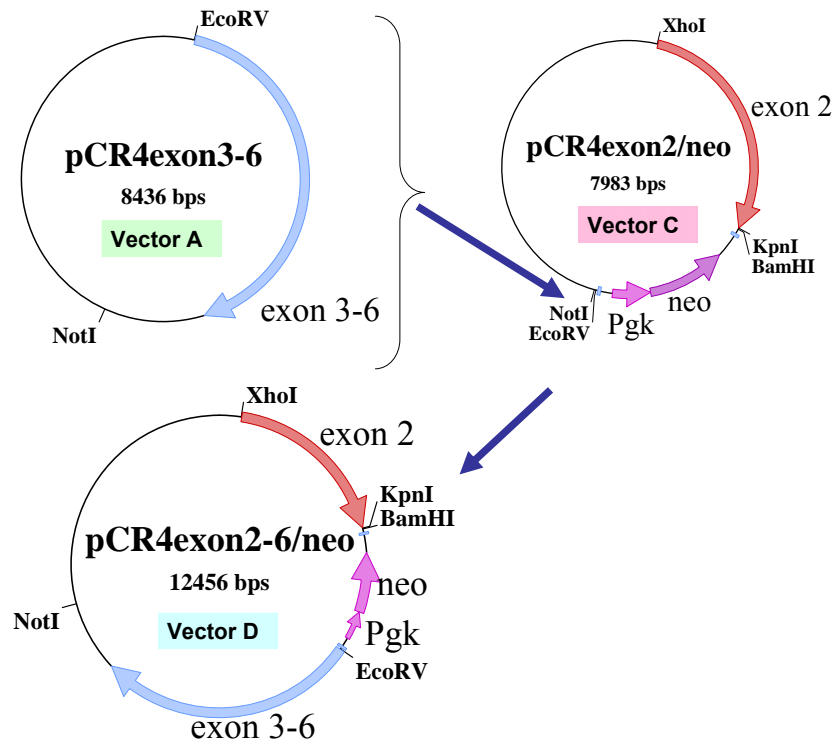


**Fig 4.6 Targeting construct version 1**

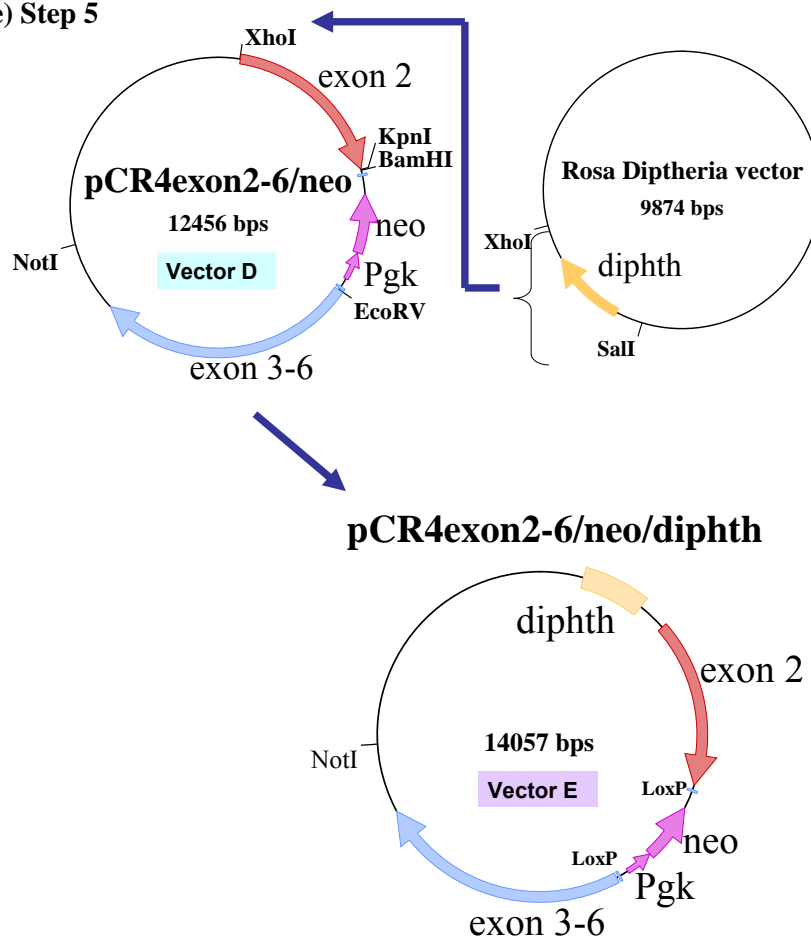
(a)-(c) Steps 1-3 are shown on this page. (d)-(e) Steps 4 and 5 are shown on the following page. The thin black line represents vector backbone, large coloured arrows/boxes are cloned DNA components of the targeting vector as labelled. Brackets indicate inserts used in ligation reactions. Relevant restriction enzyme sites are shown in bold. Dark blue arrows indicate sites of insertion



## (d) Step 4

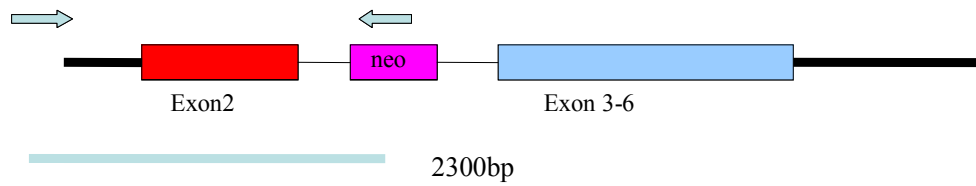
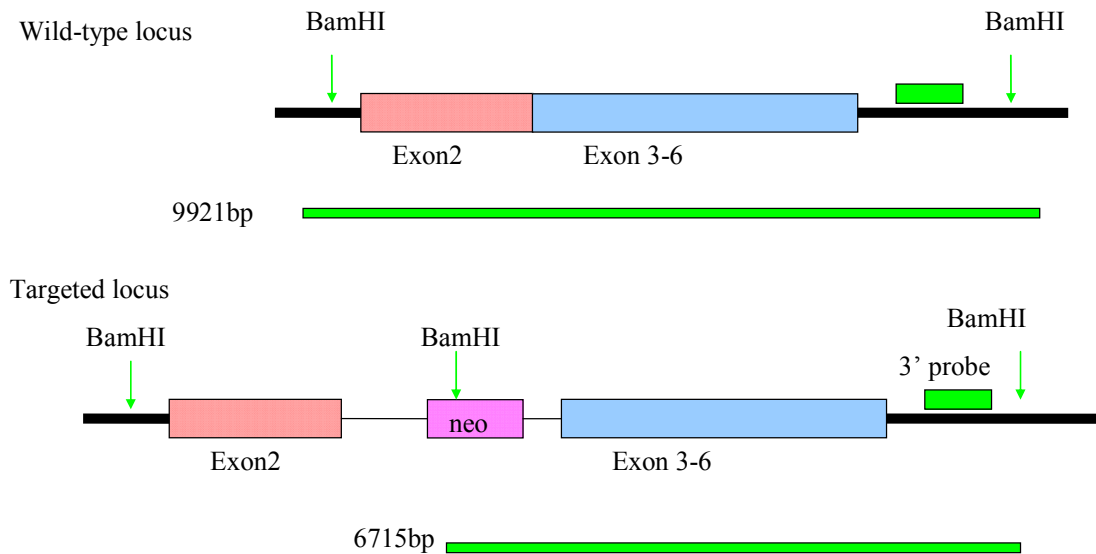
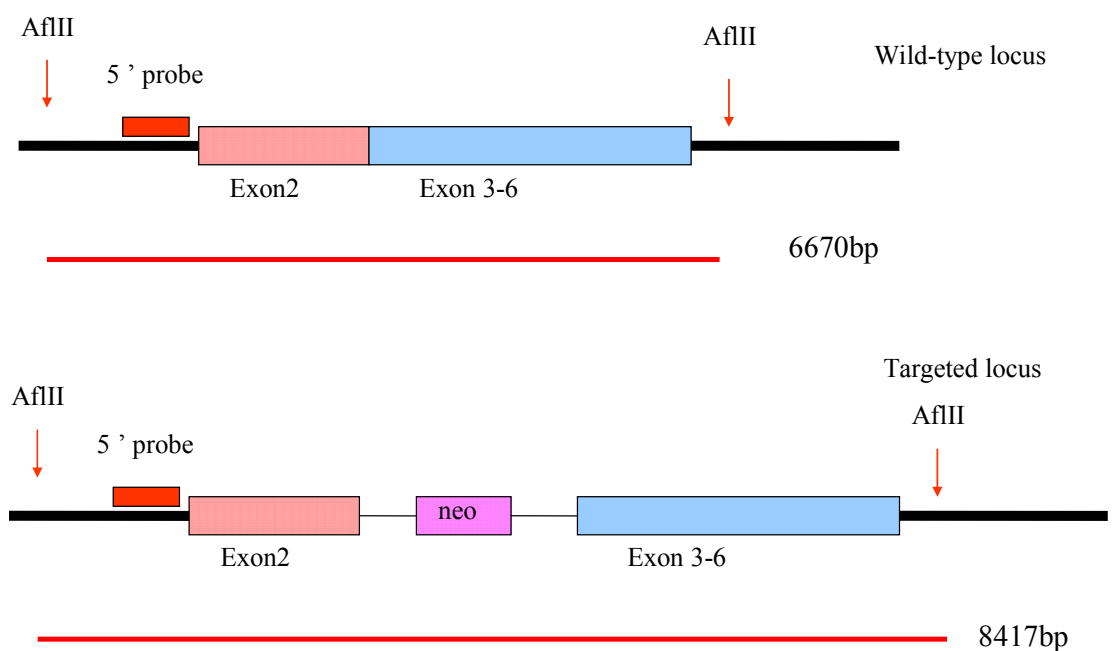


## (e) Step 5



**(a) PCR screening**

Targeted locus

**(b) Southern blot 3' probe****(c) Southern blot 5' probe****Fig 4.7 Screening strategy for targeted colonies**

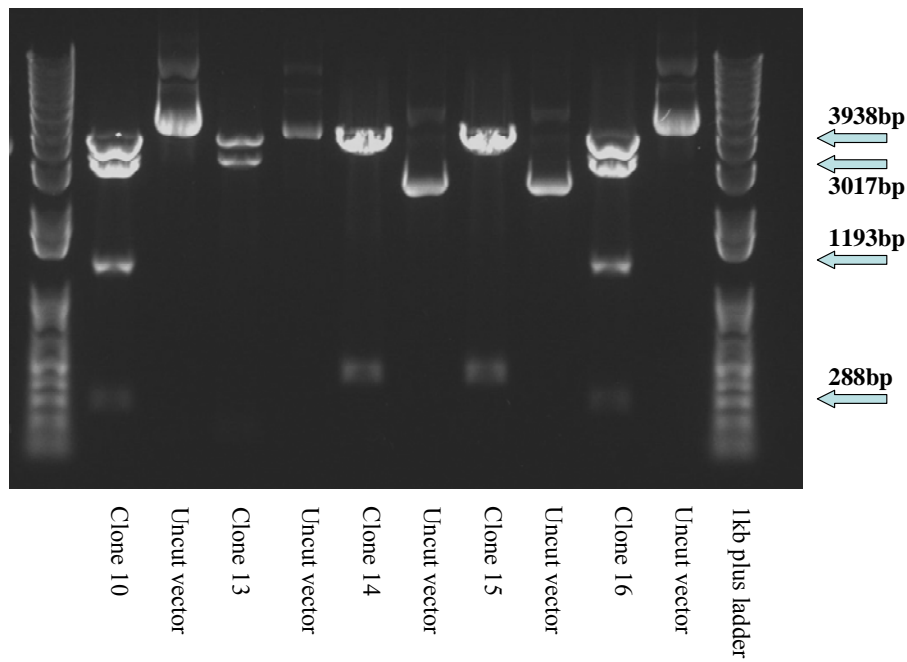
The following sections describe the individual stages involved in construction of this vector

#### **4.4.1 Vector A *pCR4 exon3-6***

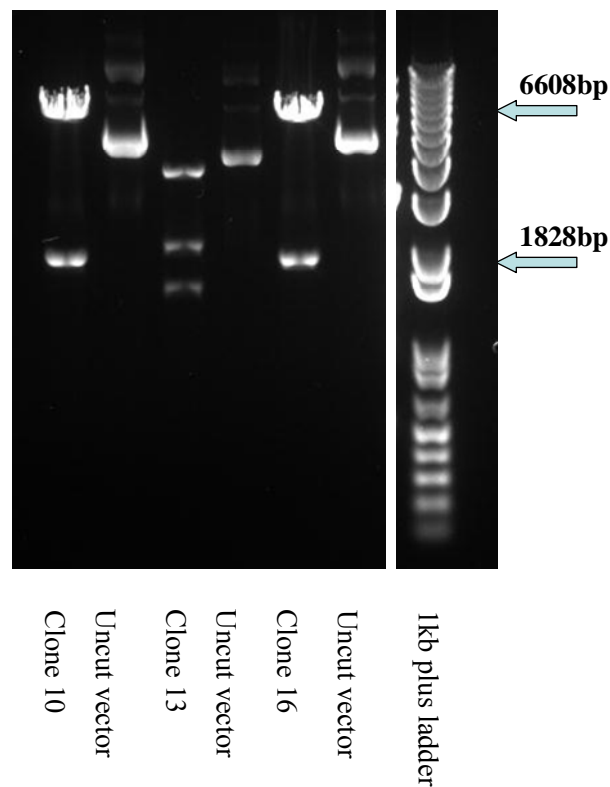
Using primers listed in table 2.10 Chapter 2, the 3' homology arm was amplified from 129 genomic DNA. An EcoRV restriction enzyme site sequence was added to the 5' end of the 5' primer to facilitate subsequent cloning steps. Use of Platinum Pfx polymerase (section 2.10.4) produced a single 4480bp product which was excised (without prior gel photography to reduce exposure to UV light), gel purified and TOPO cloned into the pCR4 vector after addition of A overhangs (section 2.12.1.1). Following overnight growth on LB-Agar plates, 5 colonies were picked for further analysis. Minipreps were made (section 2.12.2.3) and vector DNA was assessed by RE digests (section 2.7.3). The results of these digests are shown in Fig 4.8a and b. Initially an EcoRI digest was performed to assess the presence of the correct insert (fig 4.8a). The predicted pattern of fragments was seen in two of the clones (10 and 16). To assess the orientation of these inserts an NcoI digest was performed (fig 4.8b) which suggested that both clone 10 and 16 carried the correct insert in the correct orientation.

Clone 10 was sent for commercial sequencing (Agowa.de). Primer walking was used to sequence the entire 4480bp of the insert (data not shown). This confirmed 100% sequence identity with the available GATA-1 sequence (derived from the C57/Bl6 strain, GeneID 14460, locus tag MGI: 95661, Entrez Gene) with the exception of a run of 'A's in the middle of intron 4. The published sequence had 17 'A's, and clone 10 had 16'A's. This was likely to be a strain difference between 129 and C57/Bl6. If this assumption was correct then this difference would have no impact on targeting efficiency (and since it occurs in an intron, it would also have no impact on the final protein product). If it was an error introduced by the PCR amplification process (unlikely as this type of error is rarely produced by Pfu polymerases) then it was felt that this single base pair deletion was unlikely to drastically impact on targeting efficiency. Therefore, clone 10 was designated vector A and used in all subsequent cloning steps.

(a)



(b)

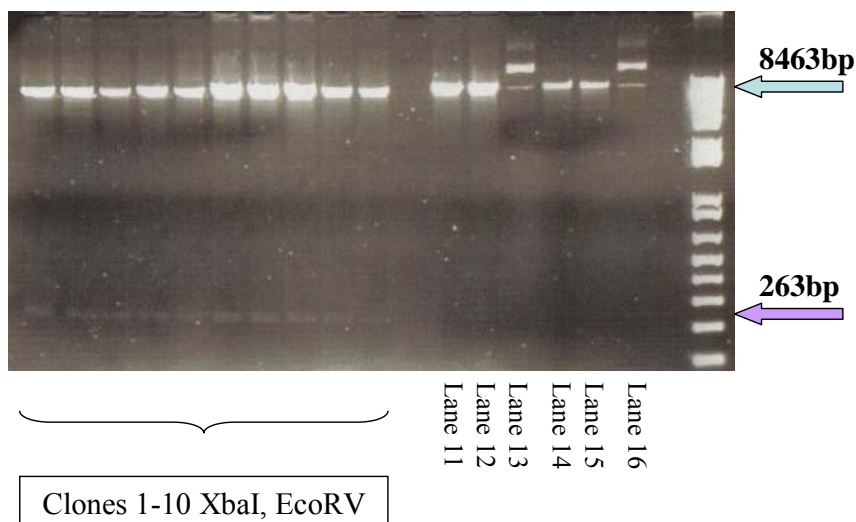


**Fig 4.8 Restriction enzyme digestion of Vector A**

(a) Results of EcoRI digest which should produce fragments of sizes 3938bp, 3017bp, 1193bp and 288bp (shown by light blue arrows) in the presence of an exon 3-6 insert. Clones 10 and 16 produce appropriate sized fragments. (b) To assess orientation an NcoI digest was performed. This should produce fragments of 6608bp and 1828bp (blue arrows) in the correct orientation or 5506bp and 2930bp in the wrong orientation. Both clones 10 and 16 produce the correct pattern on digest.

#### 4.4.1.1 Site-directed mutagenesis of Vector A

In order to produce the GATA-1FL targeting construct (i.e. capable of exclusive GATA-1FL production) it was necessary to produce a version of vector A with a mutation in the second ATG start codon thereby preventing production of GATA-1s. Primers were designed (table 2.6, chapter 2) to alter the ATG (methionine) to CTA (leucine), this 2 basepair change creates an XbaI site (which cuts at the sequence TCTAGA). Therefore successful site-directed mutagenesis can be confirmed by digestion of the mutagenised colonies with XbaI. Site-directed mutagenesis was performed using the Stratagene QuikchangeXL commercial kit (section 2.12.1.3). 10 clones were picked for further analysis. An EcoRV & XbaI double digest of mini-prepped DNA from these clones showed successful site-directed mutagenesis in all 10 as shown in Fig 4.9. EcoRV cuts once at the vector/insert junction (this site was introduced by inclusion of its target sequence in the 5'primer during initial PCR amplification, described in 4.4.1 above, to allow subsequent cloning steps) and therefore should produce a linear fragment of 8463bp (as shown by the positive control clone 10 wild-type lane 14). XbaI does not cut in the wild type molecule and therefore should produce a pattern identical to uncut plasmid DNA: lanes 13 and 16 respectively). In the presence of a mutated second ATG an XbaI site is created producing a linear 8463bp fragment when the DNA is digested with XbaI (lane 11) and two fragments of 8200bp and 263bp when a double digest using EcoRV and XbaI is performed (lanes 1-10). One of these clones (clone 6, lane 6) was used subsequently to produce the GATA-FL version of the targeting vector.



**Fig 4.9 Site-directed mutagenesis of Vector A**

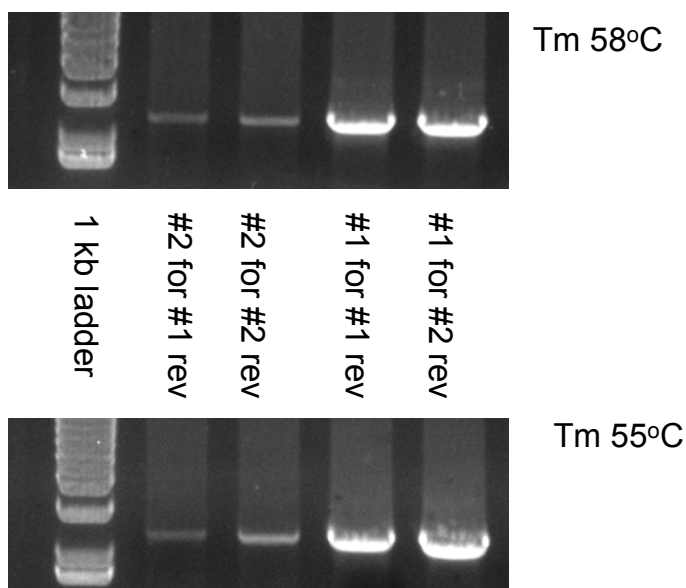
Blue arrow denotes size of linearised vector DNA, purple arrow 263bp XbaI fragment

#### **4.4.2 Vector B pCR4 exon2**

In parallel with the creation of vector A described above attempts were also made to amplify and clone the shorter 5' homology arm (2344bp) to produce vector B. Primer sequences are shown in table 2.10, chapter 2. XhoI and KpnI restriction enzyme site sequences were added to the 5' and 3' primers respectively to allow subsequent cloning steps. Initial PCR amplification from 129 genomic DNA produced a variety of products of different sizes, one of which roughly equated to the expected 2344bp fragment. This band was excised, gel purified, 'A' overhangs were added and it was TOPO cloned into PCR4. Unfortunately diagnostic digests with a variety of restriction enzymes (EcoRI, XhoI/NotI) did not produce the expected pattern of fragments (data not shown). Two further attempts were made to amplify up the 2344bp fragment with these primers, altering the annealing temperature, extension times and buffer conditions; again a fragment of approximately 2344bp was amplified on each occasion. These fragments were TOPO cloned, and produced plentiful numbers of colonies, but diagnostic digests (EcoRI, NcoI) continued to show incorrect fragment sizes (data not shown). In an attempt to narrow down colonies that had incorporated GATA-1 sequences, PCR screening was performed directly on colonies picked from the Agar plate. This involved picking colonies with a sterile pipette tip, dipping the tip into a PCR tube (containing master mix and BFT4 FL GATA 5' primer and Exon2-XhoI 3' primers (listed in table 2.5 and 2.10, Chapter 2, respectively)) and then ejecting the tip into 3ml of LB-broth with ampicillin in a correspondingly labelled universal container. Following PCR screening any promising colonies could then be grown up from their corresponding universal container culture and mini-prepped for further evaluation. 41 colonies were screened by this approach using BFT4 FL GATA forward and Exon2-XhoI reverse primers, both of which should anneal within the insert. 24 colonies produced PCR products of the correct size, but again these colonies did not produce correct fragment sizes on diagnostic restriction enzyme digests (EcoRI, NcoI, DraI, PstI) (data not shown). To investigate the origin of these abnormal inserts, sequencing was performed on a couple of the more equivocal inserts (whose fragments on RE digest roughly equated with predicted patterns). Results showed that one was derived from a mouse chromosome 7 segment and the other a mouse chromosome 2 segment. An initial screen of the primer sequences using the NCBI BLAST programme (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to look for non-specific binding to other regions of the genome had appeared satisfactory. However, on more detailed analysis, the addition of the XhoI linker sequence to the 5' primer had produced a short section of homology with this chromosome 2 sequence, and this same primer was also able to anneal to the chromosome 2 reverse strand (in the

opposite orientation) thereby allowing amplification of this segment of DNA. Amplification of the chromosome 7 segment remained unexplained.

To try to prevent this non-specific annealing of the forward and reverse primers, a second set of primers were designed (exon 2 version 2 (#2), table 2.10, Chapter 2). Since Intron 1-Exon 2 appeared to be a problematical PCR target the amplification reaction was optimised using (i) different combinations of primers (e.g. #2 forward, #1 reverse, #2 forward #2 reverse etc) (ii) two different annealing temperatures and (iii) different concentrations of Pfx amplification buffer and enhancer solutions (2.5x buffer, 2x buffer 1x enhancer, 2x buffer, 1x buffer). Results of (i) and (ii) are shown in Fig 4.10 and show that #1 forward #2 reverse gives the most intense band and that an annealing temperature of 55°C gives the highest yield. Results of (iii) i.e. optimisation of buffer conditions, cannot be shown, as bands were excised without photography, but 2x buffer with 1x enhancer gave clearly superior results to all other combinations. Use of these conditions was adopted for all subsequent amplification reactions.

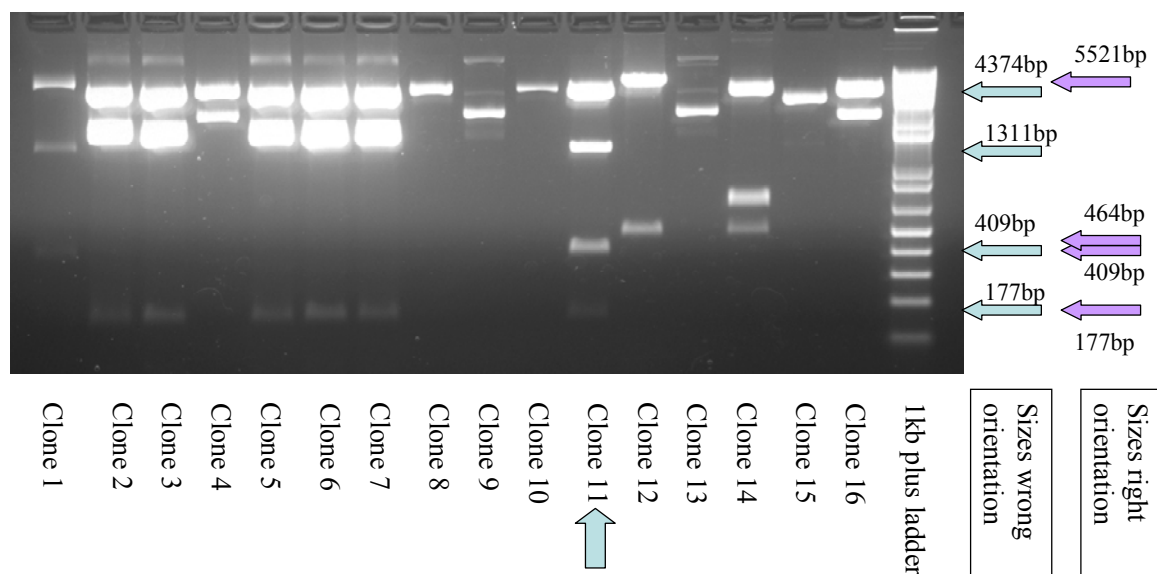


**Fig 4.10 Optimisation of exon 2 amplification**

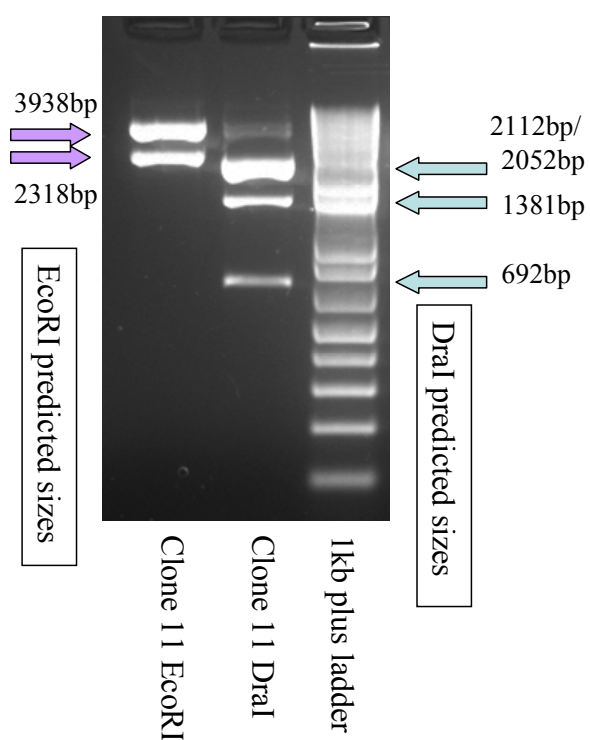
Using these new amplification conditions a single band of approximately 2300bp was excised, TOPO cloned and 16 colonies mini-prepped. To look for the presence of the insert in the correct orientation a PstI digest was performed. This should produce fragments of 5221, 464,409 and 177bp in the correct orientation and 4374, 1311, 409 and 177bp in the reverse orientation. Clone 11 appeared to produce fragment sizes corresponding to the correct insert but in reverse orientation (the presence of a 1311bp band is clearly seen as

indicated in Fig 4.11a), the identity of the insert was further confirmed by EcoRI and DraI digests (fig 4.11b) which both produced the expected size of fragments for exon 2 reverse orientation.

(a)



(b)

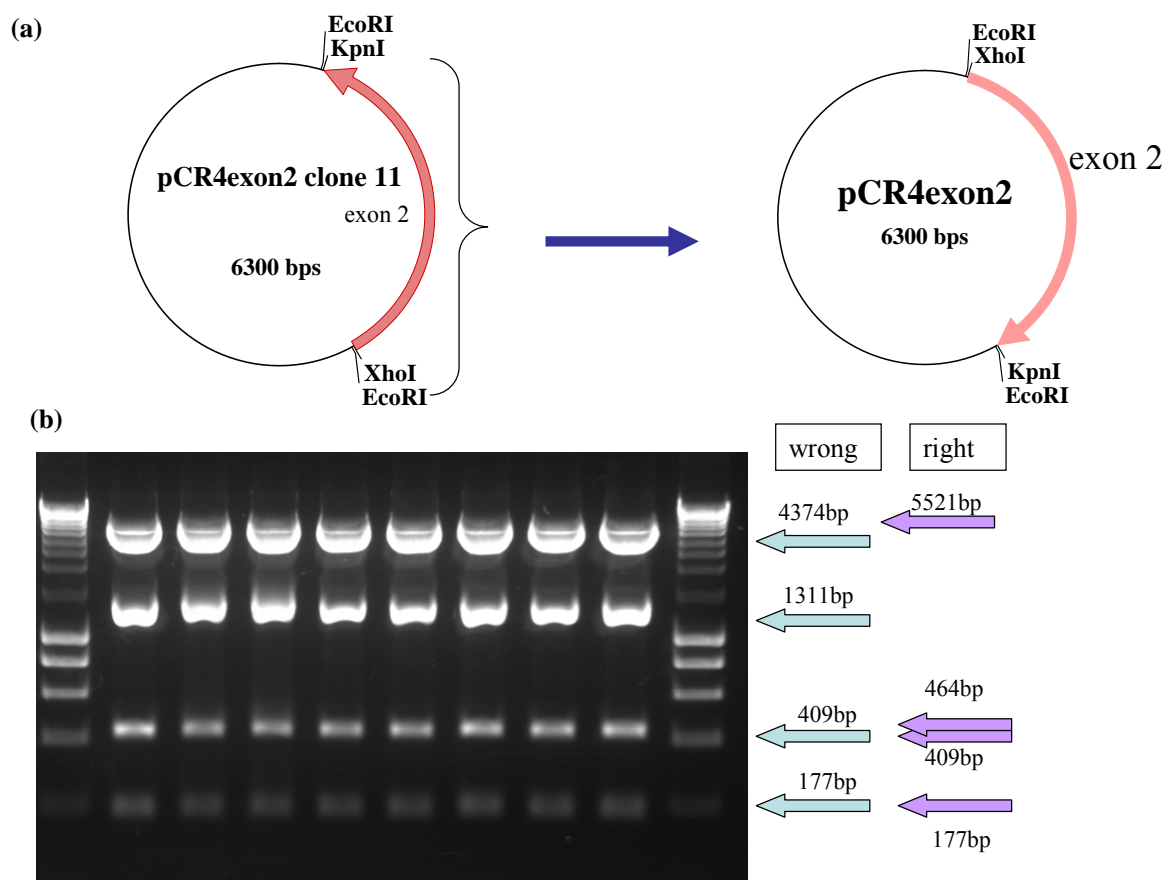


**Fig 4.11 Restriction enzyme digestion of clone 11 vector B**

(a) PstI digest produces fragment sizes compatible with an exon 2 insert but in the incorrect orientation in clone 11. Expected fragment sizes for the right orientation are shown in purple, wrong orientation in blue. (b) EcoRI (purple arrows) and DraI (blue arrows) digests confirm the expected fragment sizes for exon 2 reverse orientation in clone 11.



Having eventually cloned the exon 2 PCR fragment it was decided to try and reverse its orientation by exploiting the *EcoRI* restriction enzyme sites in the 5' and 3' multiple cloning sites as shown in Fig 4.12a, rather than attempt further cloning of primary PCR products. 5µg of clone 11 plasmid DNA were digested with *EcoRI*. The vector (3938bp) and insert (2318bp) bands were excised and gel purified. The vector was dephosphorylated (as described in 2.12.1.2) to prevent religation of the compatible sticky ends without recombining with the insert. Ligation was then performed using standard conditions (2.12.1.2). Control reactions consisting of vector only and insert only were run to ensure that colonies produced on the test plate were due to religation between vector and insert, rather than contamination of the ligation reaction with uncut clone 11 plasmid DNA. These plates showed very small numbers of colonies (<5) whereas the test plate had more than 200 colonies. Minipreps of 8 colonies showed all 8 had religated in the incorrect orientation (fig 4.12b). A further 24 colonies were picked and again *PstI* digest confirmed the reverse orientation for all inserts (data not shown). Religation in this orientation would be predicted to occur in only 50% of the colonies.



**Fig 4.12 Attempt to reverse orientation of clone 11**

At this point it was decided to attempt an alternative cloning strategy as the cloning of exon 2 in the correct orientation appeared technically very difficult. This may be due to secondary structure within this region of DNA or other unknown factors. Results of the alternative strategy are described below.

## 4.5 Targeting construct version 2

Following the inability to clone vector B (pCR4exon 2) in correct orientation, a second strategy was devised. This is illustrated in Fig 4.13, in summary:

**Step 1:** Transfer exon 2 pCR4 (clone 11- incorrect orientation), into pBluescript II KS-cloning vector using XhoI/EcoRI digest of pCR4exon2 and XhoI/EcoRI digest of pBSII to produce vector 1 (pBSIIexon2) Fig 4.13a.

*Site-directed mutagenesis was then performed on this construct to delete the first ATG (GATA-1s targeting construct)*

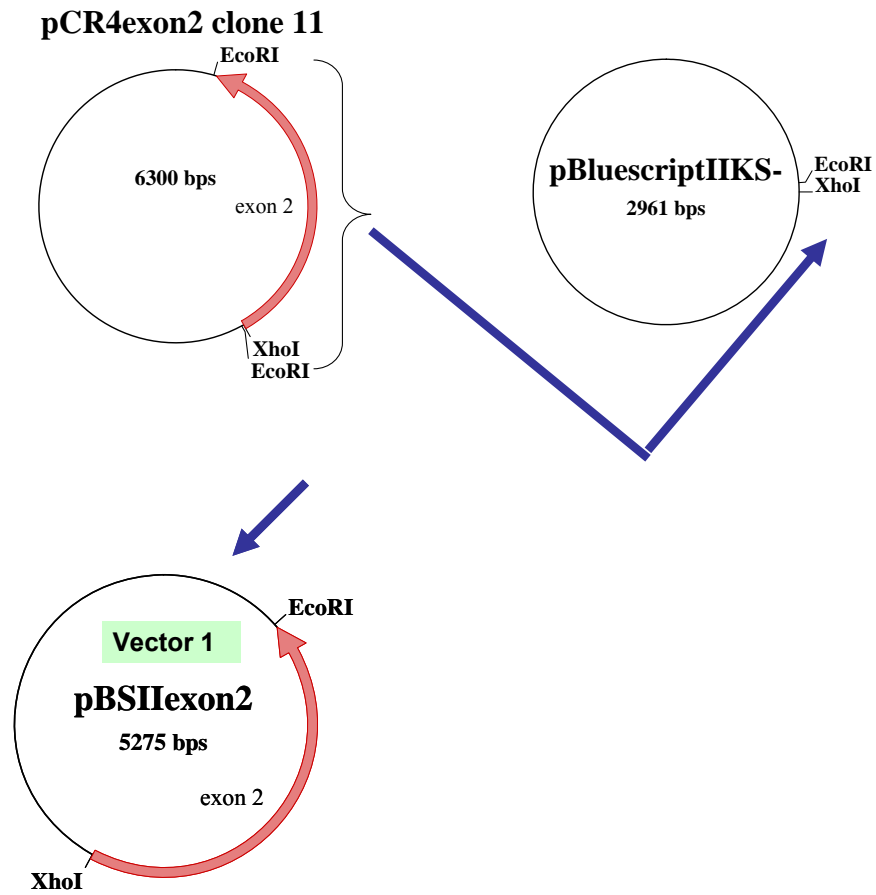
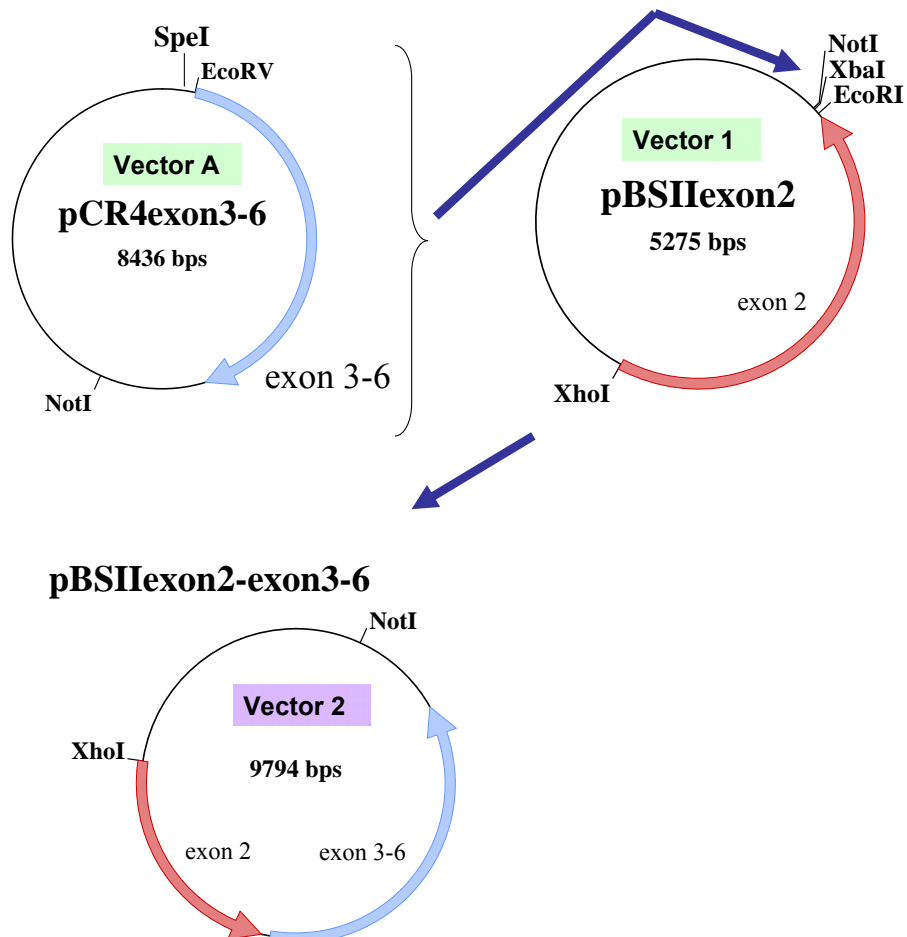
**Step 2:** Ligate exon3-6 into pBSIIexon2 (vector 1) by NotI/SpeI digest of pCR4exon3-6 (vector A) and NotI/XbaI digest of vector 1 (XbaI and SpeI produce compatible sticky ends), to produce vector 2 (pBSIIexon2-exon3-6). Fig 4.13b.

**Step 3:** Ligate pgk-neo into vector 2 using BamHI/EcoRV digest of pgkneo and BamHI/SmaI digest of vector 2 (EcoRV and SmaI are both blunt-ended cutters), to produce vector 3 (pBSIIexon2-pgkneo-exon3-6). Fig 4.13c.

**Step 4:** Ligate in diphtheria toxin via XhoI digest of vector 3 and XhoI/SalI digest of Rosa diphtheria toxin vector to produce vector 4 (pBSIIexon2-pgkneo-exon3-6-diphth). Fig 4.13d.

This final targeting construct could be linearised with NotI prior to electroporation

The screening strategy to identify targeted clones is essentially the same as that shown in Fig 4.7 with the exception that the neomycin selection gene is in reverse orientation requiring the use of alternative 3' primers for PCR screening as listed in table 2.10, chapter 2. Because of this reverse orientation the position of the neomycin BamHI restriction site is also altered with respect to the 3' probe, resulting in a 5095bp targeted fragment instead of a 6715bp fragment (compared to a 9921bp wild-type fragment).

**(a) Step 1****(b) Step 2**

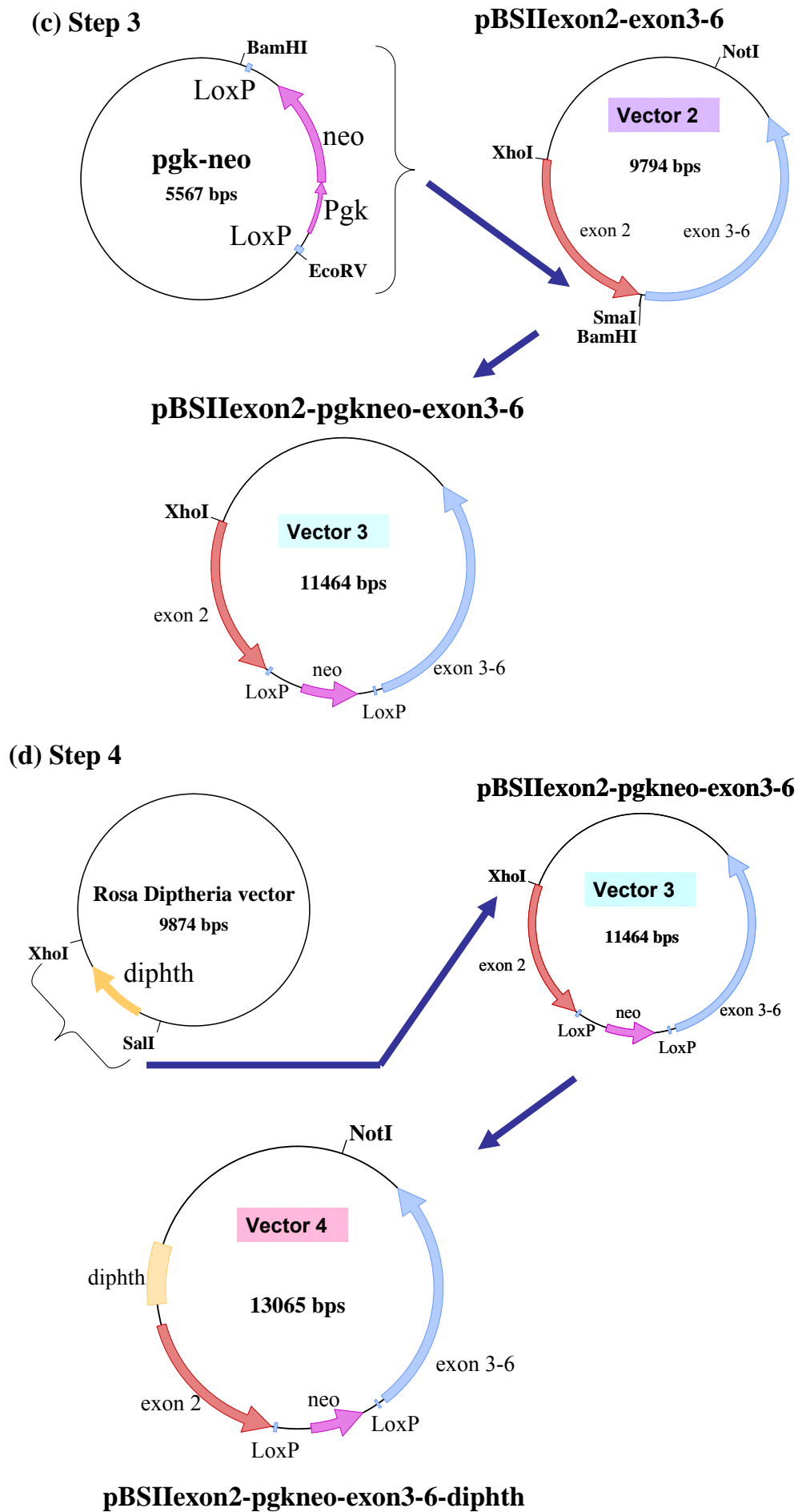


Fig 4.13 Outline of targeting strategy 2

### **4.5.1 Step 1 cloning of vector 1 (pBSII exon 2)**

Clone 11 (section 4.4.2 above) i.e. exon2 PCR4 in the wrong orientation, was sequenced to confirm that it matched the published GATA-1 genomic sequence. Like the exon 3-6 clone, a single stretch of repetitive intronic sequence (on this occasion a string of GACT repeats) showed variation between the amplified 129 genomic DNA and the published C57Bl6 sequence (6 repeats versus 12 repeats respectively) again this was ascribed to inter-strain variation, as this kind of error is not usually due to aberrant DNA polymerase activity.

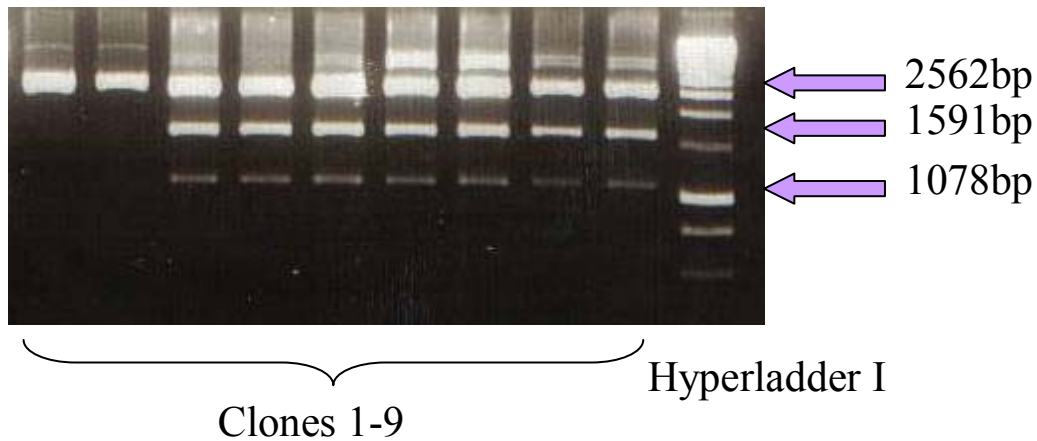
5µg of clone 11 and 5µg of the cloning vector pBluescriptIIKS- were digested with XhoI & EcoRI (using compatible buffer – REact2 (Invitrogen)). The insert (clone 11) 2303bp band, and the vector (pBSII) 2928bp band were excised, gel purified and used in a standard DNA ligation reaction (2.12.1.2) before transformation into DH5α “library efficiency” cells and growth overnight in LB broth with Ampicillin. Again control reactions (vector only, insert only) were set up to ensure there was no contamination with uncut clone 11 or pBSII vectors. Several thousand colonies were produced on the test plate and only occasional colonies seen on the control plates. 6 colonies were picked and mini-prepped. Initial digest with XhoI/EcoRI confirmed the presence of insert with the expected 2303bp and 2928bp fragments on digest (data not shown). One of these clones (clone 2) was picked for further analysis. To double check that these colonies did not represent contaminating clone 11 plasmid DNA a second set of restriction enzymes – ScaI/XbaI- were chosen that only cut in the insert and pBSII. Again the predicted size of fragments was confirmed 2562bp, 1557bp and 1112bp (versus 3246bp and 3010bp in clone 11) (data not shown).

Clone 2 was used as vector 1 in all subsequent steps.

#### **4.5.1.1 Site-directed mutagenesis of vector 1**

To produce a targeting construct capable of exclusive GATA-1s production a mutation in the first ATG was engineered to prevent synthesis of GATA-1FL but still permit use of the second ATG in exon 3 codon 84. Primers were designed (table 2.6, chapter 2) to alter the ATG (methionine) to ATC (isoleucine), this single basepair change abolishes an NcoI site (which cuts at the sequence CCATGG), and creates a ClaI site (which cuts at the sequence ATCGAT) as shown in Fig 4.5. Site-directed mutagenesis was performed using the commercial Quikchange SDM XL kit (Stratagene) as described in 2.12.1.3. Nine colonies were picked and screened for successful mutagenesis using a ScaI/ClaI double digest

(using the compatible buffer React 6) which should produce fragments of 2562bp, 1591bp and 1078bp in the presence of the introduced *Cla*I site, this expected pattern was seen in 7/9 colonies (fig 4.14). Clone 9 was picked for use in creation of the GATA-1s targeting construct.



**Fig 4.14 Site-directed mutagenesis of vector 1**

*Sca*I/*Cla*I digestion of vector 1. Purple arrows denote expected fragment sizes in the presence of an introduced *Cla*I site. Successful mutagenesis is seen in clones 3-9.

#### **4.5.2 Step 2 cloning of vector 2 (*pBSIlexon2-exon3-6*)**

Step 2 involved combining the 5' and 3' homology arms into a single targeting vector. At the end of this process two different targeting vectors were produced i.e.

1. Combination of exon 2 wild-type (vector 1) with exon 3-6 GATA-1 FL (vector A, product of site directed mutagenesis described in section 4.4.1.1 above) produced the GATA-1FL targeting vector
2. Combination of exon 2 GATA-1s (vector 1, product of site-directed mutagenesis described in 4.5.1.1 above) and wild-type exon 3-6 (vector A) produced the GATA-1s targeting vector.

5µg of vector 1 (wild-type or GATA-1s) DNA were digested with *Not*I/*Xba*I using a sequential digest i.e. digestion with *Not*I for 1 hour using REact 3 buffer at 37°C followed by purification using a Qiagen PCR purification kit (as described in section 2.7.5) and subsequent use in a *Xba*I digest using REact 2 buffer. 5µg of insert vector A (GATA-1FL or wild-type) were digested with a sequential *Not*I/*Spe*I digest using the same approach

except for use of React4 buffer in place of React2 for the second digest. It should be noted that XbaI and SpeI produce compatible sticky ends allowing subsequent ligation of these products. Following restriction enzyme digestion the products were gel purified and used in a standard ligation reaction (section 2.12.1.2). On this occasion only very sparse small colonies were produced which may reflect the increasing size of the vector (now 9794bp in size). 14 GATA-FL colonies and 3 GATA-1s colonies were picked. Miniprep DNA was used in a diagnostic double digest with BamHI/ScaI restriction enzymes (using NEBuffer 3, New England Biolabs). Clones 7, 10 and 14 GATA-1FL produced the predicted set of fragments (data not shown). No correct GATA-1s clones were seen. A second set of RE digests using EcoRI and a double XbaI/NotI digest showed that only clone 14 contained the correct insert, the other 2 clones appear to have incorporated the vector backbone from vector A. This may be because the two vector A fragments (vector backbone and insert) were similar in size after NotI/SpeI digestion (3910bp and 4526bp respectively) and the 4526bp band may have been contaminated with small amounts of the smaller 3910bp band when the band was excised from the gel. To guard against this in subsequent ligations, a triple digest was used with NotI/SpeI and PvuI which produced fragments of 4526bp, 2175bp, 1572bp and 163bp. Not only did this ensure that the 4526bp band was more easily separated from the vector backbone but also the vector backbone (digested with PvuI) no longer had compatible SpeI/NotI sticky ends making its ligation into vector 1 impossible. This strategy was used in a second ligation reaction to obtain further GATA-1s colonies. A better yield of colonies was seen and 18 were picked for further investigation. Of these 9 produced correct sized fragments on BamHI/ScaI double digest, indicating successful ligation. Again results were checked using further combinations of enzymes using EcoRI/ClaI and EcoRI/XbaI double digests which produced the expected fragment sizes (data not shown).

Vector 2-FL (clone 14) and vector 2-Sh (clone 4) were used in subsequent cloning steps:

### **4.5.3 Step 3 introduction of *pgkneo***

The next step was to “clone-in” the neomycin phosphotransferase antibiotic resistance gene (*neo*) into the targeting vectors. A plasmid containing this gene driven by the mouse *Pgk-1* promoter and flanked by Lox-P sites was already in use in the laboratory. The Lox-P *pgk-neo* cassette was transferred into vector 2-FL and vector 2-Sh using a combination of blunt ended and sticky ended ligation. The targeting vectors were cut with BamHI and the blunt cutting enzyme SmaI whilst the *pgkneo* cassette was excised from its vector backbone using BamHI and EcoRV (EcoRV is also a blunt cutter). Both reactions required

sequential digests using BamHI buffer (proprietary NEB buffer) initially, followed by purification and a second digest with EcoRV (React2) or SmaI (React 4). SmaI digestion was performed at 25°C as this is its optimum working temperature. Ligation was performed using standard conditions and 12 colonies for each construct (FL and Sh) were picked for further evaluation. Due to the increasing size of the targeting vector (now 11464bp) minipreps were performed without the use of commercial spin columns (as described in section 2.12.2.3). An EcoRI digest was performed to look for correct incorporation of the pgkneo cassette. Clones 1, 3, 9 and 10 produced the correct pattern for vector 3-Sh and clones 16, 18, 19, 20, 21, 22 and 23 for vector 3-FL. These results, and existence of the expected mutagenesis sites, were confirmed with a second set of digests ClaI/BglII (GATA-1s) and XbaI/BglII (GATA-1FL) respectively, the correct pattern of fragments was seen in all cases (data not shown).

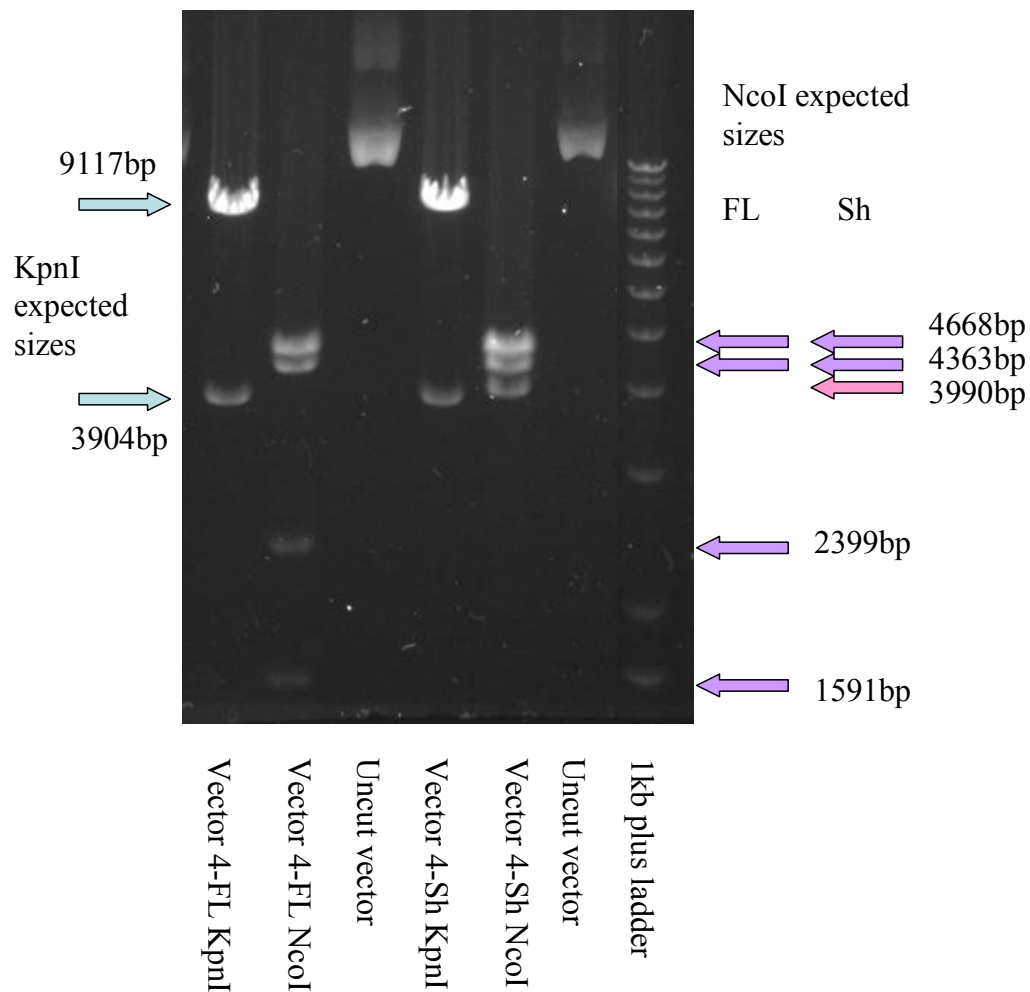
#### **4.5.4 Step 4 introduction of diphtheria toxin**

In order to enrich for colonies that have undergone homologous recombination a diphtheria toxin cassette was cloned into the 3' end of the targeting vector. The diphtheria toxin gene driven by the ubiquitously active murine Pgk-1 promoter (Zambrowicz, *et al* 1997), was available in a plasmid vector in the laboratory. The cassette could be excised using flanking XhoI/SalI restriction sites. XhoI and SalI produce compatible sticky ends and can therefore be ligated into a corresponding XhoI site in targeting vector 3. This will produce an insert in either orientation. Since the cassette contains an integral promoter it will be functional in either orientation.

5µg of vector (vector 3-FL and vector 3-sh) and insert (Rosa-diphth) plasmid DNA were digested using XhoI and XhoI/SalI respectively. The linearised vector was gel purified and then dephosphorylated to prevent vector religation. A standard ligation reaction was set up and 12 colonies picked for each targeting construct. A XhoI/BamHI digest was performed to confirm the successful introduction of the diphtheria expression cassette (data not shown). Clones 2, 5, 11 and 12 (GATA-1FL) and 13, 14, 15, 16, 18, 21 and 23 (GATA-1s) produced the expected size of fragments. These results were confirmed with additional EcoRI, NotI, and BstXI digests (data not shown) and KpnI and NcoI digests which are shown in Fig 4.15. Note that the latter NcoI digest also confirms successful abolition of the NcoI site in the GATA-1s targeting construct (i.e. a single 3990bp band (denoted by a pink arrow) in place of the wild-type 2399bp and 1591bp bands (purple arrows)).



To further check the correct construction of the targeting vectors (vector 4-FL and vector4-Sh) and orientation of the two homology arms a number of PCR reactions were performed. These used forward primers that annealed in the short homology arm and reverse primers annealing in the long homology arm so producing a product that crossed the neomycin resistance gene. These reactions produced the correct sized bands and identity was further confirmed by restriction enzyme digestion of the PCR products (data not shown).



**Fig 4.15 Diagnostic digests to confirm construction of the targeting vector**

Blue arrows denote expected sizes of KpnI digest fragments (9117bp & 3904bp). Purple/pink arrows denote expected sizes of NcoI digest fragments (4668bp, 4363bp, 2399bp & 1591bp for GATA-1FL, 4668bp, 4363bp, 3990bp (pink arrow) for GATA-1s)

## 4.6 Introduction into ES cells

Once the targeting constructs had been made (vector 4-GATA1FL and vector 4-GATA1s) they were maxi-prepped using an endotoxin-removal protocol and methods suitable for large vectors (described in 2.12.2.3). 2µg of vector DNA was linearised and a 5µl aliquot was run on a gel to ensure effective RE digestion and to confirm purity of the plasmid DNA. The remainder was ethanol-precipitated and resuspended in TE under sterile conditions. Low passage ES cell clones, known to be competent for germ-line transmission, were used for transfection. Cells were harvested and resuspended in sterile PBS and electroporated according to the protocol described in section 2.12.3.5, chapter 2. Cells were plated in ES cell media at a density of  $1 \times 10^5$ /ml (sufficient to produce 50-100 colonies per plate after selection). At 24 hours after transfection, media was replaced with selective medium containing 175µg/ml G418 and this medium was changed daily. A control plate was set up containing mock-transfected cells to monitor the effectiveness of G418 killing. As predicted, all the control cells died after approximately 5 days. Resistant colonies were sufficiently dense for picking at about day 10.

Colonies were picked as described in section 2.14.3. For the first targeting attempt, 96 GATA-1FL colonies and 132 GATA-1s colonies were picked. For attempts two and three 192 colonies were picked for each construct. All colonies were plated in duplicate with one plate used to extract genomic DNA (as described in section 2.7.1) for use in screening assays as described below and the other master plate frozen down for later recovery of any positive/equivocal colonies. Individual colonies were numbered 1-96 and 1-132 and given the prefix FL or SH to identify the targeting construct used.

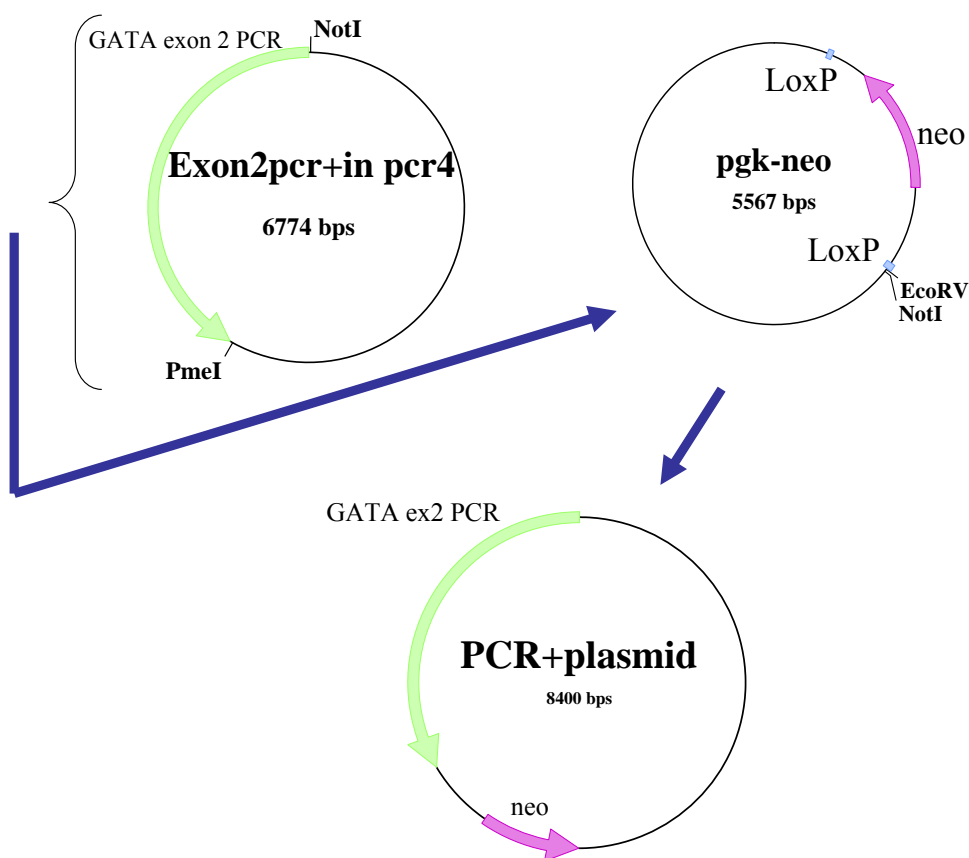
## 4.7 Screening colonies

Following extraction of genomic DNA, screening for targeting was performed in two ways. Initially all samples underwent PCR screening. However, given the low pick-up rate with this technique all samples were also screened by Southern blotting, initially using a BamHI digest with a 3' probe. A confirmatory AflIII digest with a 5' probe was performed in equivocal/positive cases. The screening protocol was performed as described in section 4.2.3 and illustrated in Fig 4.7

### 4.7.1 PCR screening

#### 4.7.1.1 Optimisation of technique

As discussed in section 4.2.3.1, in order to screen for homologous recombination events by PCR it is necessary to construct a positive control plasmid which mimics the endogenous targeted locus. The design of this vector is illustrated in fig 4.16 and described here. Initially an exon 2 fragment, comprising the short 5' homology arm and 511bp of upstream DNA sequence, was amplified from 129 genomic DNA using a PCR screening forward primer and exon 2 #2 reverse primers (listed in table 2.10, chapter 2). The resultant 2807bp amplified product was gel purified, 'A' overhangs were added and then it was TOPO cloned into the pCR4 cloning vector. Like vector B, discussed in section 4.4.2, this fragment also seemed to favour cloning in reverse orientation lending support to the theory that intrinsic physical properties of this region of DNA constrained its cloning efficiency. However for the purposes of designing a screening vector this was not a problem as long as correct orientation with respect to the neomycin cassette was maintained.



**Fig 4.16 Construction of PCR+ve control plasmid**

A PstI diagnostic digest confirmed the presence of insert in reverse orientation in 5/9 clones (data not shown). The insert from one of these clones (clone G) was ligated into the pGKneo vector using a NotI/EcoRV (sticky/blunt ended) digest of Pgkneo and a NotI/PmeI (sticky/blunt) digest of clone G. The 2820bp insert (clone G) and the linearised vector (pgkneo) were run on a 0.8% agarose gel, bands excised and gel purified and used in a standard ligation reaction. 16 colonies were picked and one of these produced the correct pattern of fragments with an EcoRI digest, this was confirmed with a subsequent BstXI digest (data not shown).

This PCR+ plasmid was used to optimise PCR screening conditions. Initially forward and reverse primers were tested with undiluted PCR+ plasmid DNA and a strong band of correct size (2.3kb) was produced. The ability of the primers to detect the targeted locus in genomic DNA isolated from targeted embryonic stem cells was then modelled. This involved calculation of the number of copies of the mouse genome (which is equivalent to the number of copies of GATA-1, since it is X-linked and therefore has one endogenous copy per genome) in 2µg of starting genomic DNA. Using Avogadro's number, the molecular weight (MW) of each basepair of DNA (approximately 650 daltons) and the number of basepairs in the diploid mouse genome ( $6 \times 10^9$ ) the number of copies of the mouse genome in 2µg of genomic DNA is calculated as follows:

Number of basepairs x MW of each basepair = MW of single copy of mouse genome

$$\text{i.e. } (6 \times 10^9) \times 650 = 3.9 \times 10^{12}$$

Copies of the genome per µl of genomic DNA (estimated concentration 2µg/µl) =

(Avogadro's number) x ( $2 \times 10^{-6}$ ) / MW of mouse genome i.e.

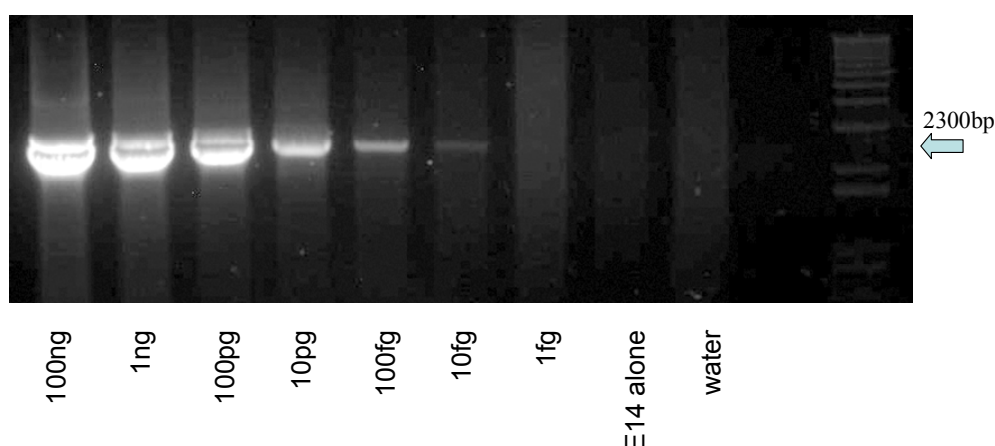
$$(6.02 \times 10^{23}) \times (2 \times 10^{-6}) / (3.9 \times 10^{12}) = (12.04 \times 10^{17}) / (3.9 \times 10^{12}) = \underline{3.08 \times 10^5 \text{ copies per } \mu\text{l}}$$

The molecular weight of PCR+ plasmid DNA was 8400bp x 650 daltons i.e  $5.4 \times 10^5$ . Therefore the number of copies of plasmid DNA (which is equivalent to the number of copies of GATA-1) in 100fg ( $1 \times 10^{-13}$ g) is:

$$(6.02 \times 10^{23}) \times (1 \times 10^{-13}) / (5.4 \times 10^5) = (6.02 \times 10^{10}) / (5.4 \times 10^5) = \underline{1.11 \times 10^5 \text{ copies per } 100\text{fg}}$$

Therefore 300fg of plasmid DNA should contain the same number of copies of GATA-1 as 2µg of genomic DNA. To test the sensitivity of the PCR primers to detect copies of

GATA-1 down at this level, serial dilutions of plasmid DNA down to 1fg/ $\mu$ l were made. However plasmid DNA is much purer and easier to amplify by PCR than crude genomic DNA isolated from ES cell lysis preparations. To add complexity 1 $\mu$ l plasmid DNA was mixed with 1.5 $\mu$ l of E14 genomic DNA before use in PCR. Results of initial titration experiments are shown in fig 4.17. These reveal detection of 10fg of template plasmid DNA. Further optimisation showed that maximum sensitivity was achieved with an annealing temperature of 55°C and 35 amplification cycles (data not shown). Therefore, using these reaction conditions, the PCR+ plasmid and primers appeared suitable for use in PCR screening of targeted clones.



**Fig 4.17 Titration of template DNA to confirm utility of PCR screening**

Blue arrow denotes expected product size, bands can be seen down to 10fg starting plasmid template. No product is seen with water or the E14 negative control DNA.

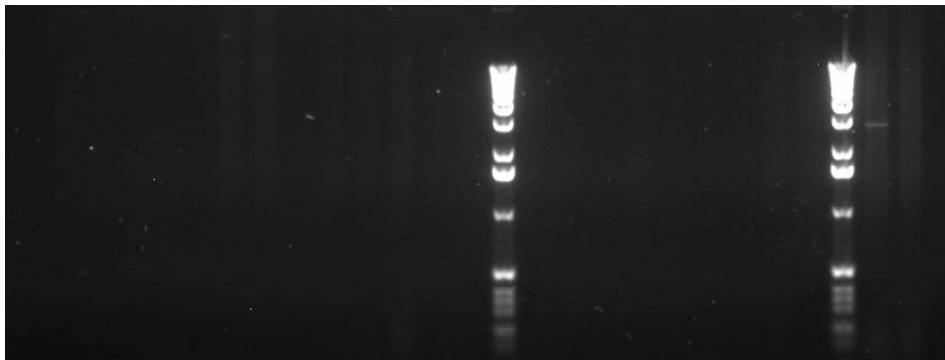
#### 4.7.1.2 Results of attempt 1

PCR screening was performed on genomic DNA extracted from all 96 GATA-1FL and 132 GATA-1s colonies. 2.5 $\mu$ l of test DNA was used in each PCR reaction and positive (PCR+ plasmid- 100fg and 10fg) and negative (H<sub>2</sub>O) template controls were run with each batch. Precautions were taken to prevent cross-contamination: PCR reactions were set up in a laminar flow hood whenever possible, separate dedicated pipettes were used for samples and the PCR+ control plasmid reactions, separate bench spaces were used to set up test samples and the positive controls. The PCR+ control was positive at the 100fg level in all batches but not always positive at the 10fg level despite identical PCR conditions to those used for initial titration (Fig 4.17). No band corresponding to the targeted allele was seen in any of the 228 samples tested – representative gels are shown in fig 4.18.

This could mean that there were no targeted alleles in any of these samples (i.e. the targeting construct had a targeting efficiency of less than 1 in 96) or that the PCR method of screening was insufficiently sensitive given the quality and quantity of the genomic DNA prepared from these relatively crude large scale extraction methods. In order to address these issues:

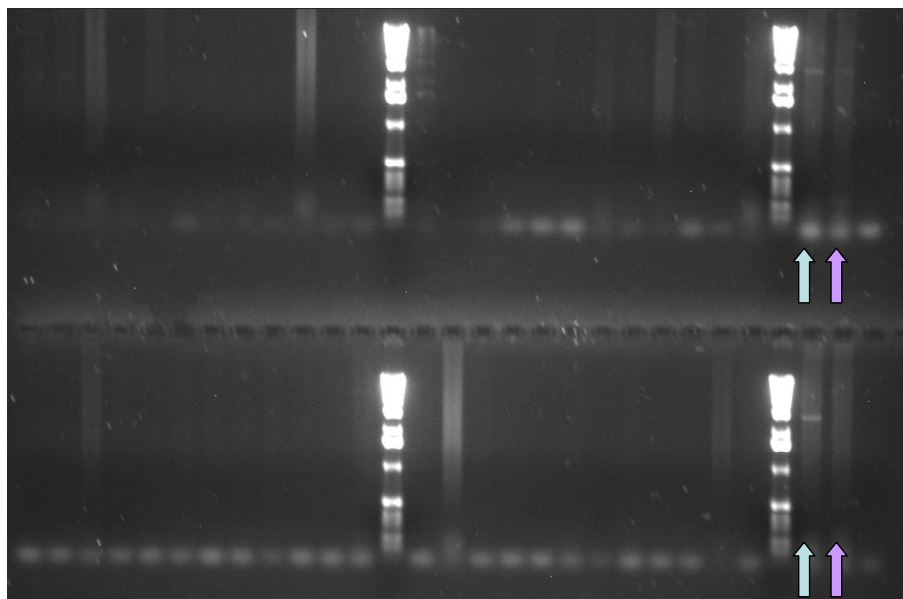
1. Colonies were also screened by Southern blotting, a tried and tested technique for identifying targeted clones under similar conditions.
2. The PCR screening strategy was made more sensitive.
3. A new batch of ES cells was targeted and more colonies picked to ensure that sufficient numbers of colonies were screened to pick up rare targeting events.

Clones FL 25-45 Targeting attempt 1



Clones SH 25-72 Targeting attempt 1

100fg ↑↑ 10fg



**Fig 4.18 PCR screening of targeting attempt 1**

### **4.7.2 Southern blotting**

Southern blotting was performed using a 3' probe and a BamHI digest as shown in fig 4.7 (although note this figure gives fragment sizes appropriate for the initial targeting vector). The neomycin cassette introduces a BamHI site into the endogenous locus and the 3' probe sits outside the targeting construct and therefore cannot pick up randomly inserted targeting construct DNA. Wild type DNA produces a 9921bp fragment whilst the targeted locus would produce a 5095bp fragment due to cleavage within the neo gene. To confirm effective targeting, a second probe which anneals to the 5' end of the endogenous locus was also used. This probe, in combination with an AflII digest produces a 6670bp fragment with wild-type DNA and an 8417bp fragment with targeting of the endogenous locus (as shown in Fig 4.7).

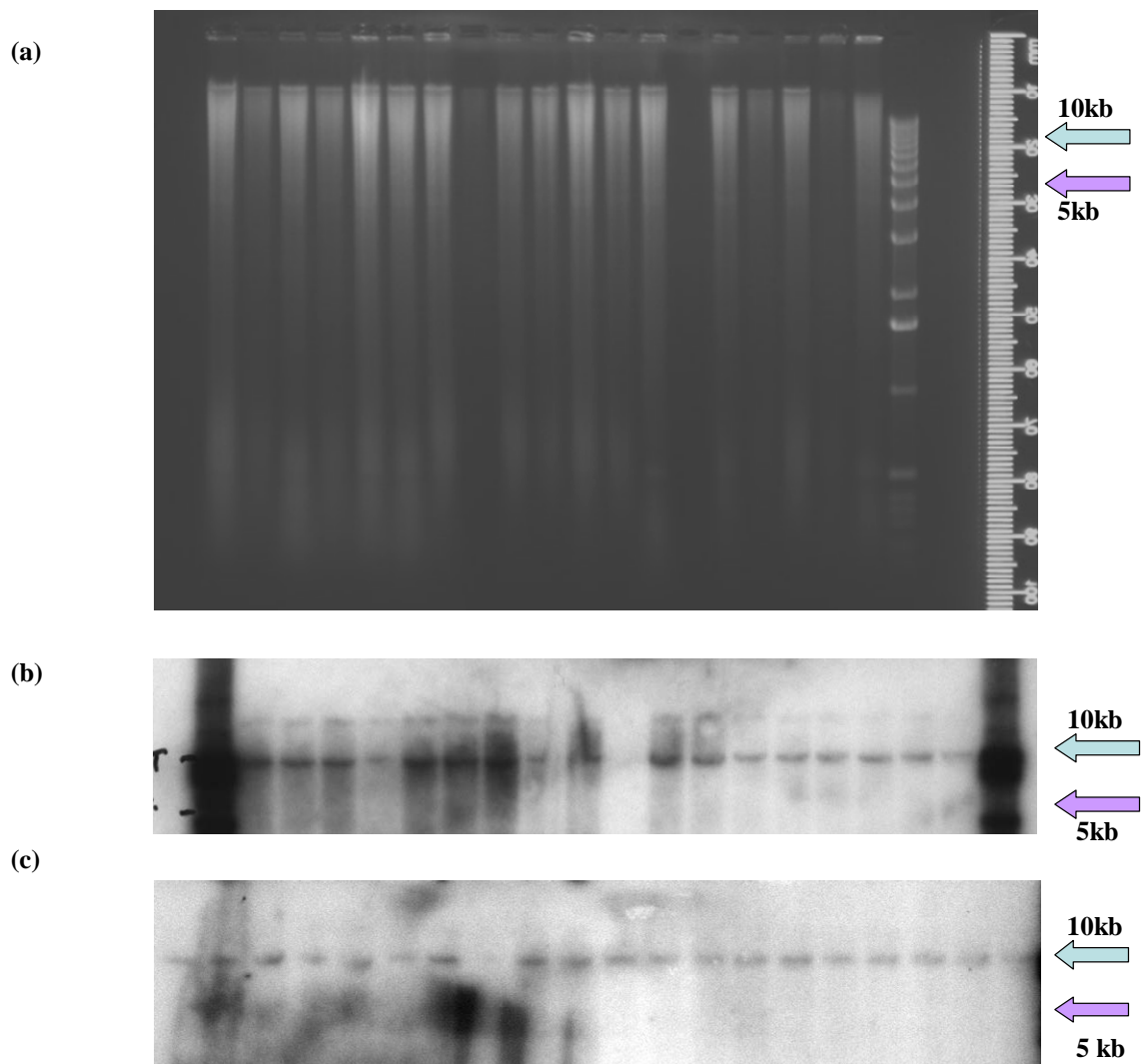
#### **4.7.2.1 Preparation of 3' and 5' probes**

The 3' and 5' probes were designed to be between 250 and 500bp long and to sit just outside the region of DNA present in the targeting constructs. PCR primer sequences are listed in table 2.10. Probes were amplified using standard PCR (2.10.1). The 5' probe was 316bp in size and the 3' probe 404bp, these products were TOPO cloned into the pCR4 cloning vector and maxi-prepped. Identity was confirmed by commercial sequencing (Agowa.de). The probes could be released from the vector using a NotI/PstI digest (NEB buffer 3) for the 5' probe and an EcoRI digest (REact3) for the 3'probe. Using 5µg of starting plasmid DNA the digestion products were gel purified and eluted in 50µl elution buffer (EB) before labelling with <sup>32</sup>P as described in section 2.7.7.4.

#### **4.7.2.2 Results of Southern blotting**

Southern blotting was performed as described in section 2.7.7. All samples were initially screened following BamHI digestion and detection using the 3' probe. This was followed by AflII digestion and detection using the 5' probe use in promising samples. Adequate DNA loading and digestion were confirmed by inspection of the ethidium bromide stained electrophoresis gel under UV light. A representative electrophoresis gel is shown in 4.19a and X-ray film detection of hybridised probe is shown in Fig 4.19b. As seen from 4.19a not all samples had adequate DNA content to allow screening (lanes) but the vast majority of samples produced a clearly identifiable wild type band (fig 4.19b and c). Since GATA-1 is X-linked and E14 ES cells are XY the presence of a wild-type band acts as a negative

control excluding targeting at the endogenous locus. All 96 GATA-1FL and 132 GATA-1s samples were screened by this method, no targeted bands were seen.



**Fig 4.19 Southern blotting of targeting attempt 1**

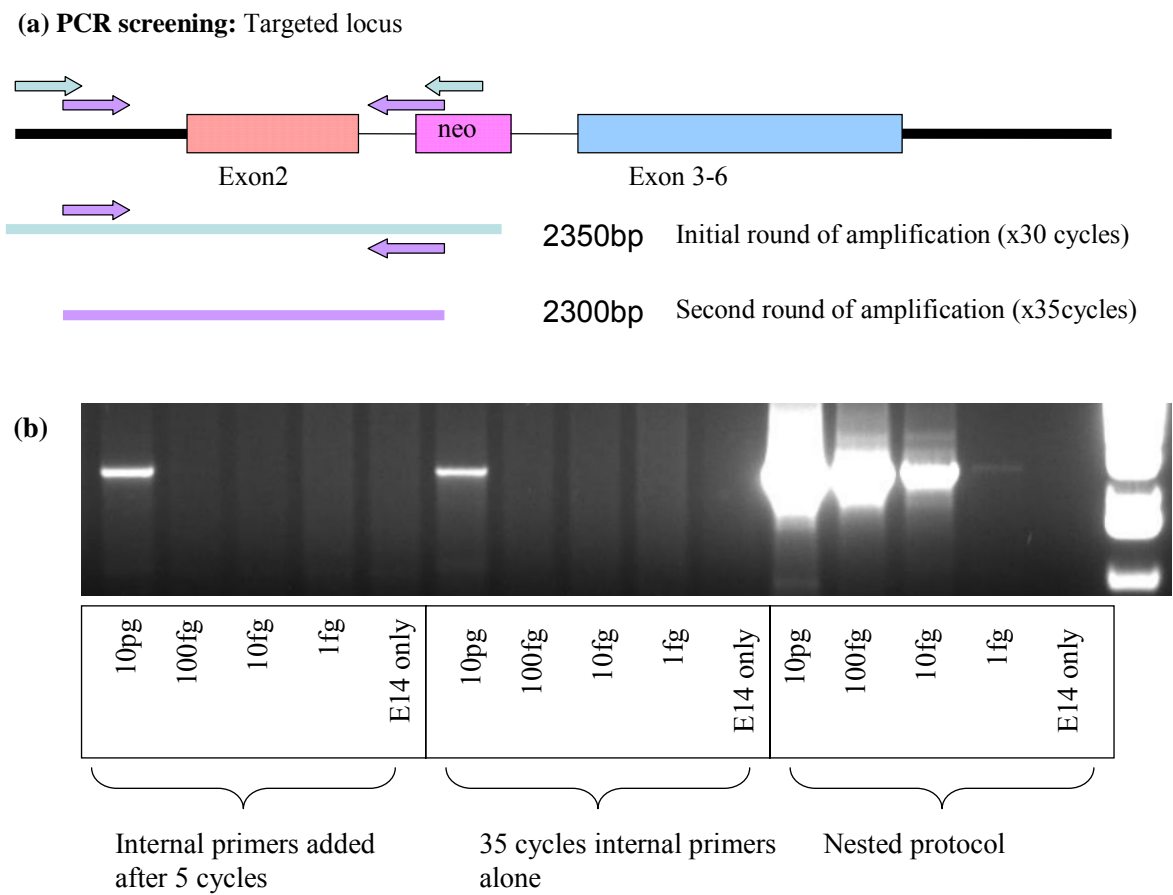
(a) Representative electrophoresis gel stained with ethidium bromide and photographed under UV illumination. Note variable DNA loading in some lanes although the majority of samples were able to produce good quality bands on hybridisation with labelled probe. (b) shows a representative radiograph (target attempt 2 FL 39-56) following 3'probe hybridisation (note (a) and (b) are not derived from the same experiment). (c) Another representative gel – target attempt 2 SH 75-94. The blue arrow shows the position of the 10kb size marker (wild-type band = 9921bp), the purple arrow the position of the 5kb size marker (targeted band = 5095bp).

### **4.7.3 Increasing the sensitivity of PCR screening**

A number of approaches were taken to try and increase the sensitivity of PCR screening. Unfortunately, use of 40 cycles of PCR amplification produced less distinct product bands on gel electrophoresis (data not shown). A commercial high yield DNA polymerase



(Phusion high fidelity DNA polymerase (New England Biolabs)) also failed to produce superior results (data not shown). Finally, new PCR primers were designed for nested PCR with external and internal primers as described in section 2.10.3. This strategy was much more effective with detectable product using 1fg of template DNA as shown in Fig 4.20. This technique was employed for all subsequent PCR screening.



**Fig 4.20 Nested PCR screening strategy**

**(a)** Diagrammatic representation of nested PCR strategy. 5µl of PCR product from the initial PCR reaction (external primers, 30 cycles) was added to a fresh PCR tube containing internal primers and a second round of amplification performed (35 cycles). **(b)** representative products after use of limited nesting protocol (addition of internal primers after 5 cycles to the same PCR tube), internal primers alone or a full nested protocol as described in (a) and section 2.10.3

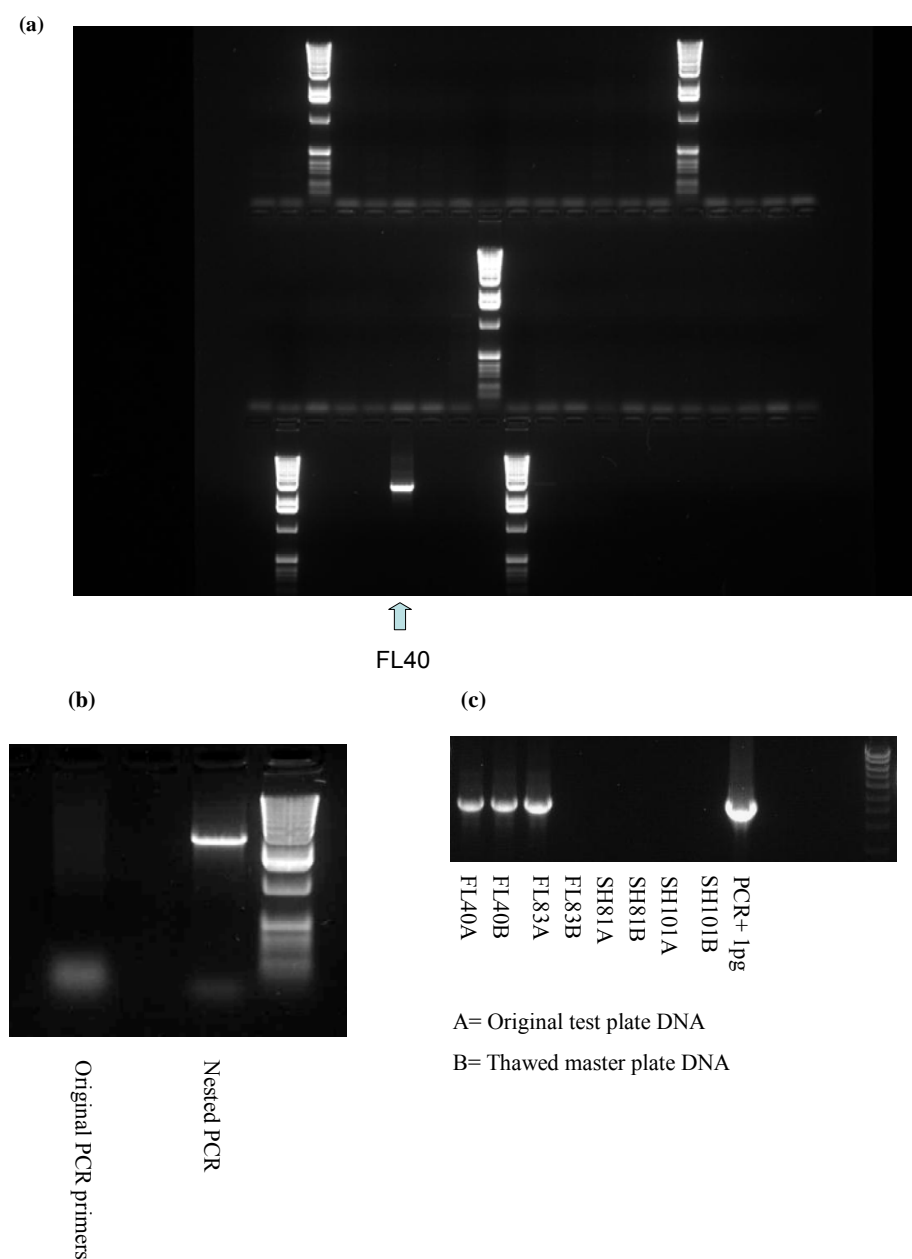
### 4.8 Second and third targeting attempts

The results above suggest that the targeting efficiencies of the two constructs are low (less than 1 in 100). Therefore a larger number of colonies need to be screened to pick up rare targeting events. A further two attempts were made to successfully target the endogenous

GATA-1 allele. In both these cases 192 colonies for each targeting construct were picked and screened by PCR and Southern blotting.

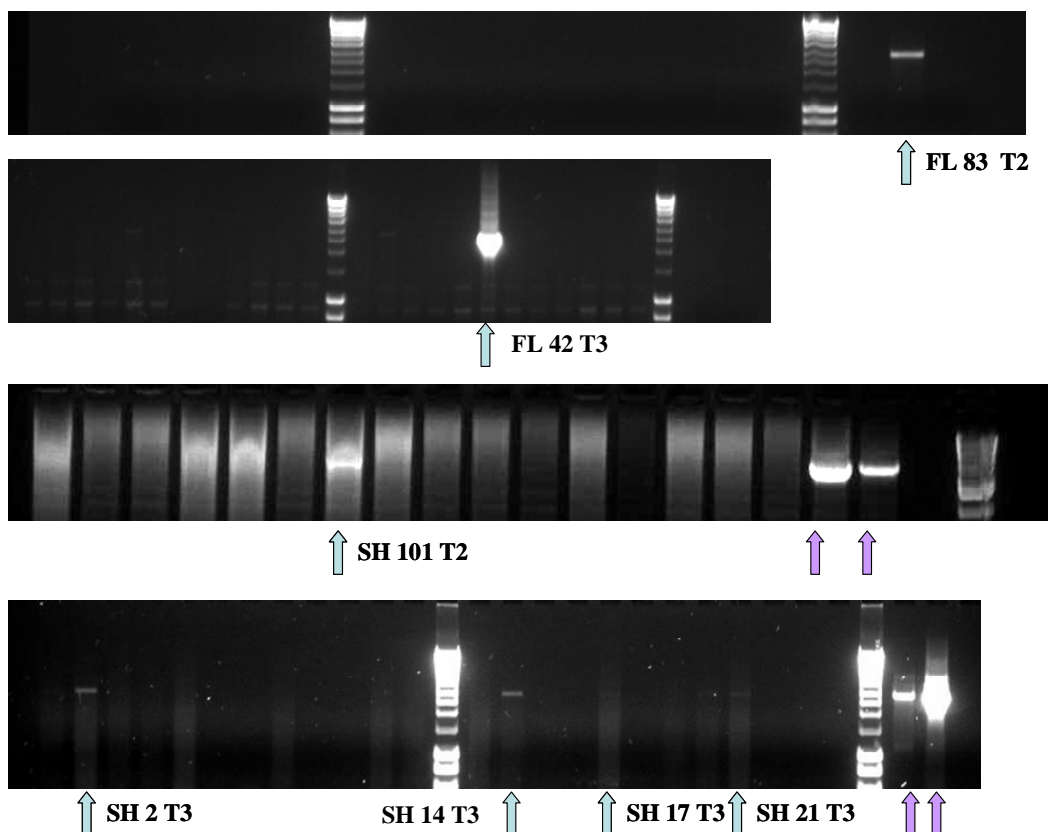
### 4.8.1 Screening by nested PCR

The use of the nested PCR protocol resulted in a number of positive bands being obtained. Fig 4.21a shows a representative positive band seen with the FL40 clone from the second targeting attempt. Following this positive result the original PCR screening strategy and the new nested strategy could be compared: Fig 4.21b shows that the FL40 sample band would not have been obtained with the original screening strategy.



**Fig 4.21 Nested PCR results for targeting attempts 1 and 2**

Therefore it remained possible that the low targeting efficiency seen in the first targeting attempt was due to inadequate sensitivity of the screening technique. To address this, all samples from targeting attempt one were re-screened with the new protocol, two of these samples - SH44 and SH81- were found to be positive with this new screen. Overall, 20 samples from the three combined targeting attempts produced positive bands by PCR as summarised in table 4.1. Some representative PCR gels are reproduced in Fig 4.22. Only 3 of these showed positive bands on an independent repeat PCR using the original DNA. To ensure the band was not due to contamination of the original DNA sample with the positive control plasmid, the corresponding master plates were then thawed (as described in 2.14.3.1). Following recovery of the cells, master stocks were frozen down in 1.5ml cryovials (as described in section 2.5.4) and remaining cells underwent cell lysis and extraction of genomic DNA as previously described (section 2.7.1). This fresh DNA was used for repeat PCR screening and Southern blotting. Repeat PCR on this new DNA was still positive for the FL40 sample as shown in Fig 4.21c but was negative for a number of other samples including SH81 (target attempt 1), FL83 (target attempt 2) and SH101 (target attempt 2) as shown in fig 4.21c.



**Fig 4.22 Positive PCR screening results from T2 and T3**

Nested PCR was performed as described in text. Blue arrows show lanes with positive bands - clone numbers and targeting attempt (T2 or T3) are shown alongside. Purple arrows show +ve control plasmid lanes (100fg and 10fg template).

### **4.8.2 Screening by Southern blot**

Southern blotting was performed for all samples using BamHI digestion and hybridisation with the 3' probe. Unexpectedly the FL40 clone produced a wild-type band on Southern blot using both 3' and 5' screening strategies (the 3' BamHI blot is shown in Fig 4.24). However, 8 samples from targeting attempts 2 and 3 did produce bands in the position of the targeted allele (5095bp) (listed in table 4.1 and representative samples shown in fig 4.23). In all cases these targeted bands were accompanied by stronger wild-type bands as shown in fig 4.23a. This finding is unexpected as E14 ES cells are male and therefore only carry one copy of the X-chromosome. Therefore, this X-chromosome should either be targeted (producing a 5095bp targeted band) or wild-type (producing a 9921bp wild-type band) but not both! One possible explanation for this finding is that, despite careful plating and picking, the ES cell colonies picked represented mixed colonies rather than a single clonal population. It should be noted that only 3 of the colonies with targeted bands on Southern blot had positive PCR results (SH2, SH14 (target attempt 2) and SH93 (target attempt 3)). This supports the need to screen with both strategies but also suggests that some results might be artefactual.

All PCR positive samples and all samples with mixed colonies on Southern blot underwent AflII digestion and hybridisation with the 5'probe. All these blots showed the presence of wild-type DNA although the quality of these blots varied making ruling out the coexistence of an additional targeted band difficult as shown in Fig 4.23b.

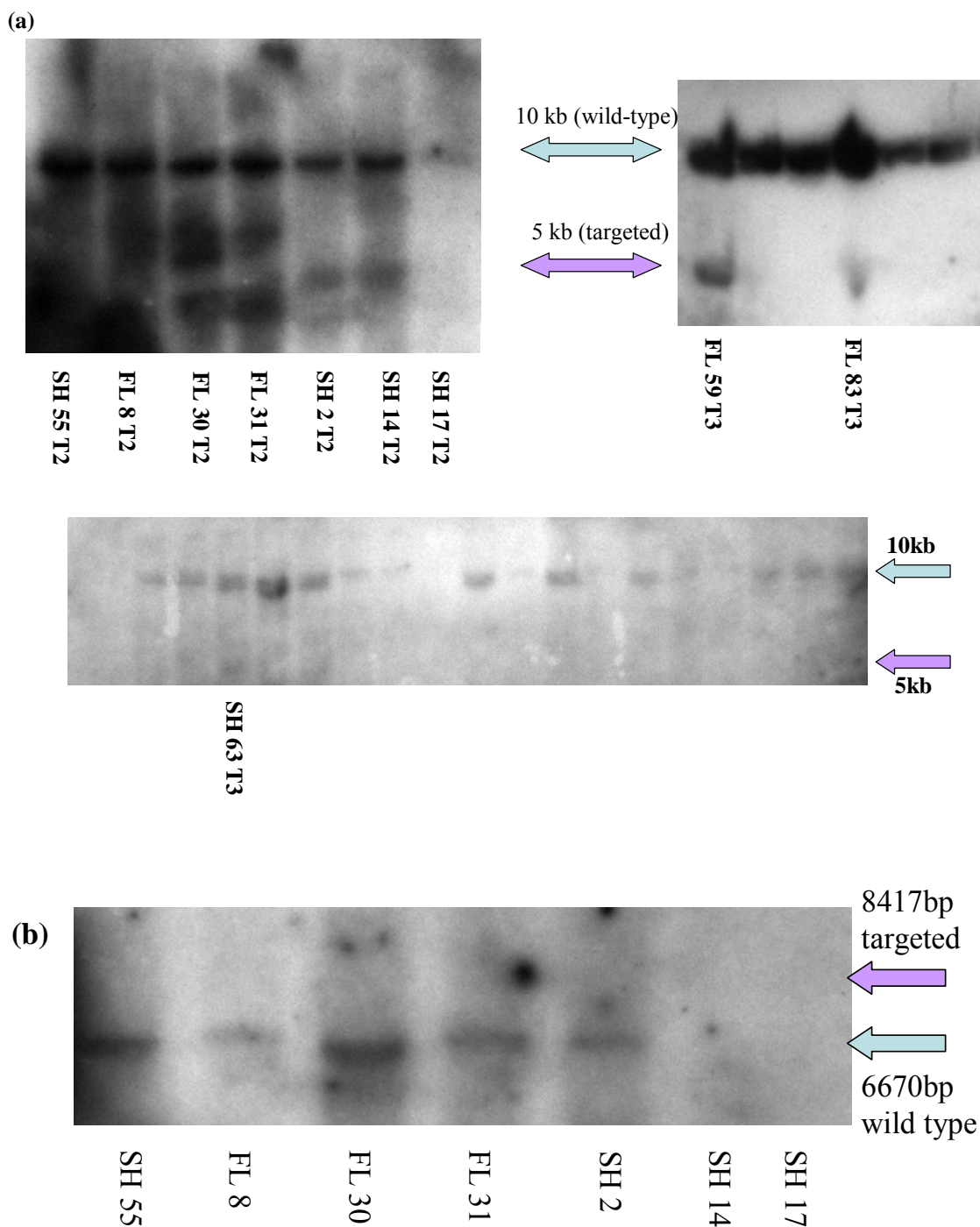
### **4.8.3 Further investigation of positive clones**

Results summarised above threw up a number of unexpected findings i.e. (i) despite targeted bands on PCR, wild type bands were seen on Southern blot (ii) targeted bands on Southern blot were accompanied by wild type bands in all cases (iii) 3'probe targeted bands were not confirmed on repeat with a 5'probe screening strategy.

Overall no clone satisfied all the criteria necessary to establish it as unequivocally targeted. There are a number of possible explanations for these findings:

1. The positive bands obtained from the original PCR could be false positives in some or all samples. This is supported by the lack of reproducibility in many cases using either the original sample or the thawed master plate. It is also supported by the increasing number of positive bands with each targeting event (by targeting attempt 3 over 800 PCR reactions

had been performed, making environmental contamination with control plasmid DNA highly likely). This is certainly possible for many of the samples but unlikely for FL40 which consistently gave high level positive results.



**Fig 4.23 Southern blotting of targeting attempts 2 and 3**

(a) Clones SH 2 & SH 14 (targeting attempt 2 (T2)) FL 59, FL 83 and SH 63 (targeting attempt 3 (T3)), appear to show co-existence of wild-type (9921bp, top blue arrow) and targeted (5095bp, bottom purple arrow) bands on BamHI digest with detection using a 3' probe. (b) these results are not confirmed on AflIII digest of the same DNA with detection using a 5' probe, in these cases only wild-type bands are clearly seen.

Target attempt	Clone	PCR	Repeat PCR	PCR master plate	S. blot (3'probe)	S. blot (5'probe)	Further action
1	SH44	+	-	ND	WT	WT	PCR band TOPO cloned and confirmed to represent targeted sequence not screening primer. Plated at low density and picked 96 subclones, all negative by PCR screening and Southern blot
1	SH81	+	-	-	WT	WT	
2	FL40	+	+	+	WT	WT	
2	FL83	+	+	-	WT	WT	Screen 24 subclones by nested PCR. Subclone 22 positive by PCR, negative by S.blot
2	SH101	+	+	-	WT	WT	
2	SH129	+	-	ND	WT	WT	
3	SH2	+	-	+	Mixed	WT	Screen 24 subclones by nested PCR. All negative
3	SH14	+	-	-	Mixed	WT	Screen 24 subclones by nested PCR. All negative
3	SH17	+	-	-	WT	WT	Screen 24 subclones by nested PCR. All negative
3	SH21	+	-	-	WT	WT	Screen 24 subclones by nested PCR. All negative
3	SH36	+	-	ND	WT	WT	Screen 24 subclones by nested PCR. All negative
3	SH63	-	ND	ND	? Mixed	WT	
3	SH93	+	-	ND	? Mixed	WT	
3	SH117	+	-	ND	WT	WT	Screen 24 subclones by nested PCR. All negative
3	FL4	+	-	ND	WT	WT	
3	FL15	+	-	ND	WT	WT	
3	FL29	+	-	-	WT	WT	Screen 24 subclones by nested PCR. All negative
3	FL38	+	-	+	WT	WT	Screen 24 subclones by nested PCR. Subclones 11 & 16 positive by PCR, negative by S.blot
3	FL42	+	-	-	WT	WT	Screen 24 subclones by nested PCR. All negative
3	FL44	-	ND	+	Mixed	WT	BamHI Southern blot remained positive (mixed) on repeat
3	FL48	-	ND	+	? Mixed	WT	Screen 24 subclones by nested PCR. Subclone 11 positive by PCR, negative by S.blot
3	FL59	-	ND	ND	? Mixed	WT	Screen 24 subclones by nested PCR. Subclones 2, 13, 19 & 22 positive by PCR, negative by S.blot
3	FL83	-	ND	ND	? Mixed	WT	
3	FL102	+	-	ND	WT	WT	

WT= wild type, ND = not done

**Table 4.1 Investigation of possible targeted clones**

2. The positive PCR bands could represent true targeted alleles at the borderline of sensitivity by PCR detection therefore not always positive on repeat. This is not supported by the Southern blot results which did not show corresponding targeted bands in many cases.
3. If the original picked colonies were mixed rather than clonal then the variable results with original test DNA and master plate DNA could represent differences in the proportion of the targeted and non-targeted clone. It is possible that the targeted clone had some kind of growth or survival disadvantage (possibly due to silencing or reduced expression of the neomycin selection gene) and therefore was out-competed when the cells were thawed a second time. This mixed colony theory also supports the co-existence of wild-type and targeted bands on Southern blot.
4. The wild-type band seen in the BamHI digest could be a composite band of GATA-1 and artefactual contaminating DNA (such as partially digested microsatellite DNA) picked up incidentally by the GATA-1 probe. Therefore the WT band seen in the presence of the targeted band could be artefactual. This is unlikely as it would be necessary to postulate that the non-specific and GATA-1 bands always ran in the same position and were therefore indistinguishable, and also that independent AflII digest also produced an artefactual band.
5. The targeted band seen in the BamHI digest could be a false positive. BamHI is a 6 basepair cutter meaning that it cuts on average every 4096 basepairs in the genome. If the GATA-1 3'probe exhibited any non-specific binding to any of these random fragments then an approximately 5kb sized band would be a relatively common occurrence. This theory would explain why targeted bands were never seen on the corresponding AflII digest as this random event would be highly unlikely to be duplicated using a different restriction enzyme and probe in the same clone. It also explains the apparent discrepancy between PCR and Southern blot results i.e. positive Southern blots were not always accompanied by positive PCR reactions (such as clones SH63, FL44, FL48, FL59 and FL83, table 4.1) and *vice versa* (such as clones SH44, SH81, FL40, FL83, SH101, SH129, etc table 4.1).
6. The targeted band in the BamHI digest could have been due to incomplete homologous recombination events. In this case the vector and the endogenous allele begin strand exchange which results in the vector picking up some of the flanking sequences of the target. However, this homologous recombination intermediate does not proceed to full

exchange and the vector dissociates from the target and may go on to randomly integrate elsewhere in the genome. This would explain the presence of both wild-type and targeted sequences in the same clone. It is interesting that mixed colonies were only ever seen with the BamHI digest which detects homologous recombination events involving the 3' long homology arm of the vector. As described above (section 4.4.1) this arm was easy to clone and had almost 100% sequence identity with GATA-1. It is possible that the shorter arm, which may have significant secondary structure, may have been particularly resistant to homologous recombination. If this was the case then this type of targeting intermediate, involving 3' but not 5' exchange, may have been commoner than full recombination.

Hypotheses 1, 3 and 4 above were tested as detailed below:

#### **4.8.3.1 Are the PCR bands false positives?**

To test whether the PCR bands represented contamination with screening plasmid DNA the FL40 band (robustly positive on multiple repeat PCRs) and the PCR+ screening plasmid bands were excised, TOPO cloned and sequenced. Construction of the screening plasmid was subtly different from the targeting construct at the exon2-neomycin junction, with an additional 6 basepairs of sequence in the targeting construct. Commercial sequencing (Agowa.de) of the two PCR products confirmed that they differed by these 6 basepairs and therefore that the FL40 product could not be due to contamination with screening plasmid DNA.

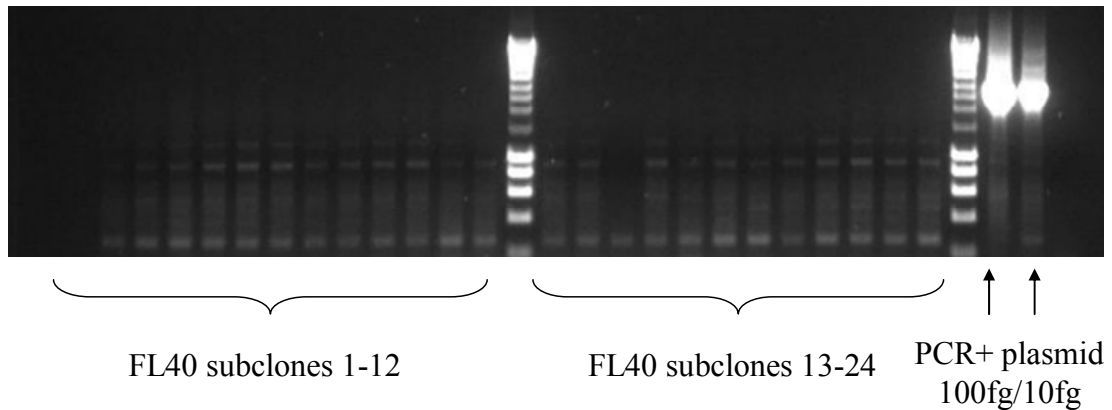
#### **4.8.3.2 Is there any evidence for mixed colonies?**

The picking of mixed colonies was felt to be the most likely explanation for the co-existence of wild-type and targeted bands on Southern blot. To address this, the FL40 sample was thawed from the frozen cryovial (described in section 4.8.1) and following cell recovery the cells were plated at low density on 10cm tissue culture plates. This was achieved by trypsinising and counting the cells, resuspending at  $1 \times 10^5$ /ml (ensuring a single cell suspension by vigorous pipetting) and then making serial dilutions to achieve plating densities of 1000 cells, 500 cells, 250 cells and 100 cells per 10cm plate. Cells were allowed to grow for 1 week and then 96 well-spaced colonies were picked for further screening, as previously described (section 2.14.3). Unfortunately, none of the 96 subclones were positive either by PCR screening or Southern blot (representative PCR gel and blot shown in Fig 4.24). It is worth noting that PCR screening was still positive in the genomic DNA extracted from the original thawing of the master plate (described in section

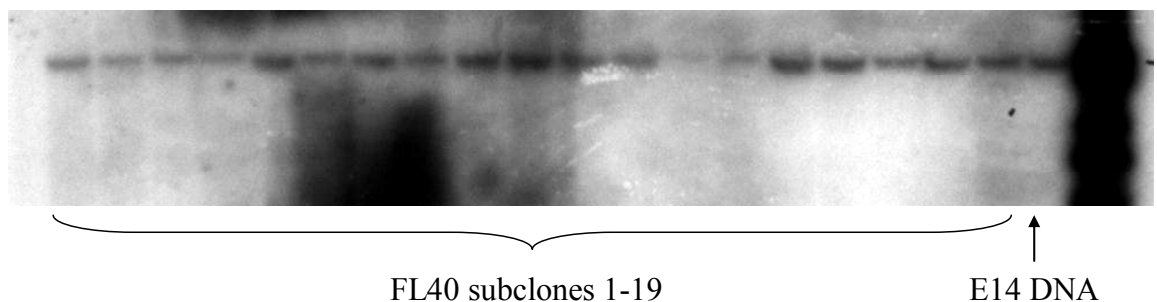


4.8.1) but not in individual colonies from this second thawing. This could either be because the targeted clone represented less than 1% of the original mixed colony (and therefore may not have been detected on individual testing of 96 subclones) or that the targeted clone had some kind of growth or survival disadvantage making its recovery on repeated freeze/thawing impossible.

### Nested PCR screen



### BamHI Southern blot

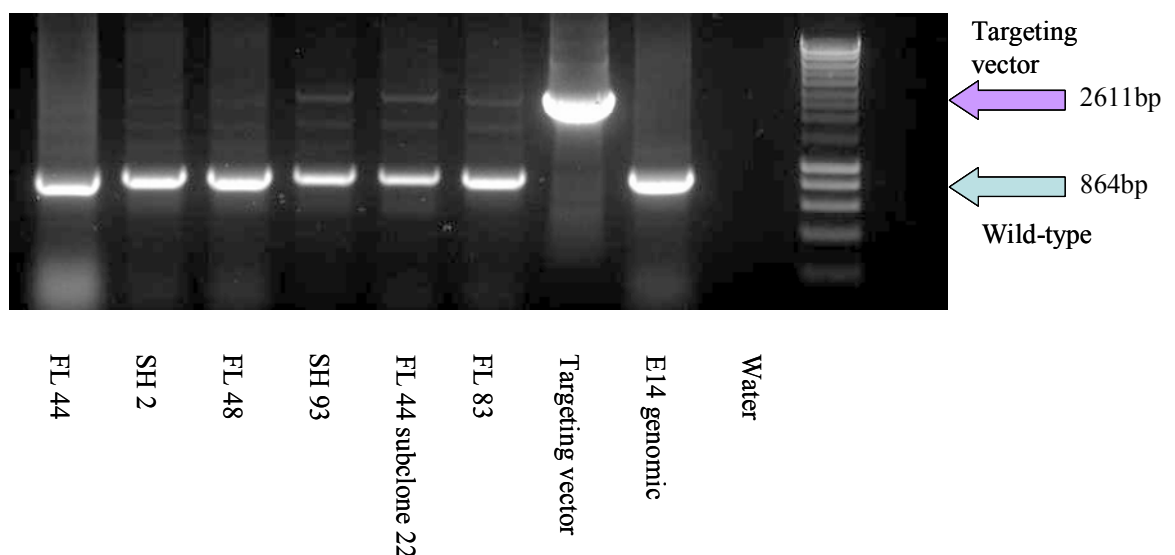


**Fig 4.24 Screening of FL40 subclones by PCR and Southern blot**

A similar approach was taken for 9 other potentially mixed colonies, either positive by Southern blot or PCR, as shown in table 4.1. These samples had not undergone previous thawing of the master plate and therefore it might have been easier to recover targeted clones. Due to practical constraints in sample handling, 24 subclones were picked for each colony, this would be expected to pick up targeted clones if they represented more than about 5% of the original mixed colony. All these subclones were screened by nested PCR and positive samples screened by Southern blot. Although 8 positive subclones from 4 different colonies were positive by PCR (listed in table 4.1) none of these proved positive by Southern blot (data not shown). It seems likely that a number of these PCR results were false positives, although this was not formally tested.

### 4.8.3.3 Do clones with a targeted band also possess wild-type GATA-1?

As discussed above the wild-type GATA-1 band, seen in addition to the targeted band on southern blots could conceivably be an artefact. In the presence of a correctly targeted GATA-1 locus no wild-type GATA-1 should be detectable. To test this hypothesis, colonies with a mixed pattern of bands by Southern blot were screened for the presence of wild-type GATA-1 by PCR. PCR primers were chosen that spanned the neomycin resistance gene with the forward primer (BFT4 FLGATA table 2.5, chapter 2) annealing in exon 2 (the short homology arm) and the reverse (alt spl 1 table 2.5, chapter 2) in exon 3 (the long homology arm). In the presence of wild type GATA-1 an 864bp product should be seen, whilst a targeted allele or randomly inserted targeting construct would produce a 2611bp product (incorporating the intervening neomycin cassette). The targeting vector was used as a positive control for introduced GATA-1 whilst untargeted native E14 DNA was used as a positive control for wild-type GATA-1, water template acted as a negative control for both reactions. The results are shown in fig 4.25. A strong wild-type GATA-1 band was seen in all colonies tested (FL44, SH2, FL48, SH93, FL83) all of which had produced a mixed pattern on southern blotting. The presence of the larger targeting construct band was also seen although much fainter, suggesting that these clones had randomly integrated the GATA-1 targeting construct.



**Fig 4.25 PCR reveals wild-type GATA in all colonies tested**

The purple arrow shows the expected product size for vector derived GATA-1 and the blue arrow for wild-type GATA-1. A wild-type band is seen in all colonies tested.

## 4.9 Summary of targeting experiments

Despite screening of more than 1000 colonies and subclones, no confirmed homologous recombination events were detected. There are a number of possible explanations for the failure to detect homologous recombination in these targeting experiments i.e.

**1. *Technical problems with construction of the targeting vector:*** many precautions were taken to ensure that the targeting vector was correctly constructed and contained no DNA mismatches that would compromise targeting efficiency. These are detailed in sections 4.4 and 4.5 above and include the use of isogenic DNA, full sequencing of the entire 6798bp (4480bp long arm and 2318bp short arm) of the two homology arms, and multiple checks of orientation by RE digest and PCR. Although even small degrees of DNA mismatch can compromise targeting efficiency it seems unlikely that the single basepair mismatch (GATA-1s) or the two-basepair mismatch (GATA-1FL) introduced by site-directed mutagenesis would be sufficient to prevent targeting, especially as the homology arms were longer than reported minimums for targeting (Hasty, *et al* 1991). One other possibility is that the negative selection strategy with diphtheria toxin was ineffective. Production of toxic diphtheria protein from this cassette was not formally tested. The absence of an effective negative selection marker would significantly increase the number of clones needed to be screened to pick up a homologous recombination event.

**2. *Technical problems with the screening strategy:*** although the initial PCR screening strategy was of borderline sensitivity for the detection of targeted clones the nested protocol showed reproducible bands using 1fg of plasmid DNA which was calculated to be more than two logs below the required level to pick up endogenous targeted GATA-1 (as described in section 4.7.1.1). Since the Southern blot protocol reliably picked up wild-type bands using a BamHI digest and 3' probe, it seems unlikely that it would be insufficiently sensitive to pick up a targeted band. The existence of the neomycin BamHI site was confirmed by BamHI digest of the targeting vector (data not shown). In addition, the presence of a wild-type band effectively ruled out targeting and so acted as a positive control. Only a few samples lacked wild-type bands and these always equated with low levels of DNA as visualised on the ethidium bromide stained electrophoresis gel.

**3. *Intrinsic properties of the endogenous GATA-1 locus:*** The extreme difficulties, detailed in section 4.4.1, in amplifying and then directionally cloning the exon 2 5' homology arm (despite it being a much shorter, and therefore a theoretically easier, segment than the corresponding 3' homology arm) suggest that intrinsic properties of this

region of DNA make it difficult to manipulate. Although on visual inspection it is not particularly GC rich, intron 1 contains large segments of repetitive sequence including multiple GATA binding sites. It may be that physical properties of this region of DNA make it particularly unsuitable for cloning and targeting. Unfortunately in order to produce a targeted mutation in exon 2 it is impossible to leave out this region (especially as placing a mutation near the end of a homology arm puts it at risk of not being incorporated during the homologous recombination process). Original descriptions of GATA-1 targeting in murine ES cells used targeting vectors with very short 5' homology arms (<1.1kb) (Pevny, *et al* 1991). Although the reason behind this was not reported it is possible that this reflected similar difficulties in cloning this region of DNA.

**4. Abnormalities in ES cells used for targeting:** again precautions were taken to use low passage germ-line competent ES cells for targeting to maximise the possibility of introducing stable genetic change. Nevertheless it is possible that these particular cells had acquired genomic instability making successful targeting less likely. ES cells rapidly acquire aneuploidy on repeated passage, with one report showing that only 20% of cells have euploid metaphases after 25 cell passages (Longo, *et al* 1997). One intriguing possibility is that these male ES cells could have acquired a second copy of the X-chromosome (i.e. making them XXY), this would explain the co-existence of wild-type and targeted bands in the same clone, although it would not explain the lack of an AflII 5' band. This possibility could be further investigated by karyotypic analysis of the targeted ES cell clones.

Following these exhaustive attempts to achieve targeted ES cell clones it was decided, due to time and practical constraints, to abandon further attempts at targeting. Instead experiments looking at over-expression of GATA-1 isoforms in the human K562 and Meg-01 cell lines, as well as ectopic expression in murine ES cells were pursued, as described in the following chapter.

## 5 A functional analysis of GATA-1 isoforms

---

### 5.1 Introduction

GATA-1 gene targeting, described in the previous chapter, is a particularly attractive approach to the study of gene function, mainly because it allows properly regulated expression of the transgene using the endogenous promoter and removes the endogenous source of GATA-1. However, the use of transgenic expression vectors randomly integrating into the host genome is a very well established approach for studying the functional consequences of gene (over)expression and has provided insights into the roles of several transcription factors as discussed in section 1.1.3.2 with examples given in table 1.2. This approach brings its own advantages, particularly the ability to use human cell lines (gene targeting and in vitro haematopoietic differentiation of human ES cells are not yet well validated techniques). It appears that the GATA-1s mutation may have different phenotypic consequences in human and murine systems, as evidenced by the murine GATA-1s transgenic mouse phenotype (Li, *et al* 2005) compared to the GATA-1s expressing human kindred (Hollanda, *et al* 2006). Therefore, the study of human cells may provide insights not gained in the murine system. In addition, the gene targeting approach provides limited opportunity to study the interaction with trisomy 21, so vital for the TAM/AMKL phenotype. Both murine and human systems will be examined in this chapter.

In order to study GATA-1 isoform function and interaction with trisomy 21 further, the individual GATA-1 isoforms were over-expressed in 3 different cell lines – murine ES cells (discussed in chapter 4), a human meg-erythroid cell line (Meg-01), and another human haematopoietic cell line with meg-erythroid differentiation capacity- K562. Meg-01 cells carry trisomy 21, and are capable of inducible in vitro differentiation along both megakaryocytic and erythroid lineages. Results using this system were compared to isoform over-expression in K562 cells, which are also capable of megakaryocytic and erythroid differentiation, but without trisomy 21. This system is discussed in more detail in section 5.5 below. This chapter starts with a description of the construction and validation of plasmid vectors capable of exclusive GATA-1FL or GATA-1s expression. It then goes on to describe utilisation of these vectors to examine the effects of GATA-1 isoform expression on:

1. The ability of murine embryonic stem cells to form haematopoietic colonies in semi-solid culture medium.
2. The biological properties of two cell lines -K562 and Meg-01- with meg-erythroid potential, one of which (Meg-01) is trisomic for chromosome 21.
3. Gene expression profiles, expression of cell surface markers and development of increased DNA ploidy during in vitro differentiation of K562 and Meg-01 along the erythroid and megakaryocytic lineages.

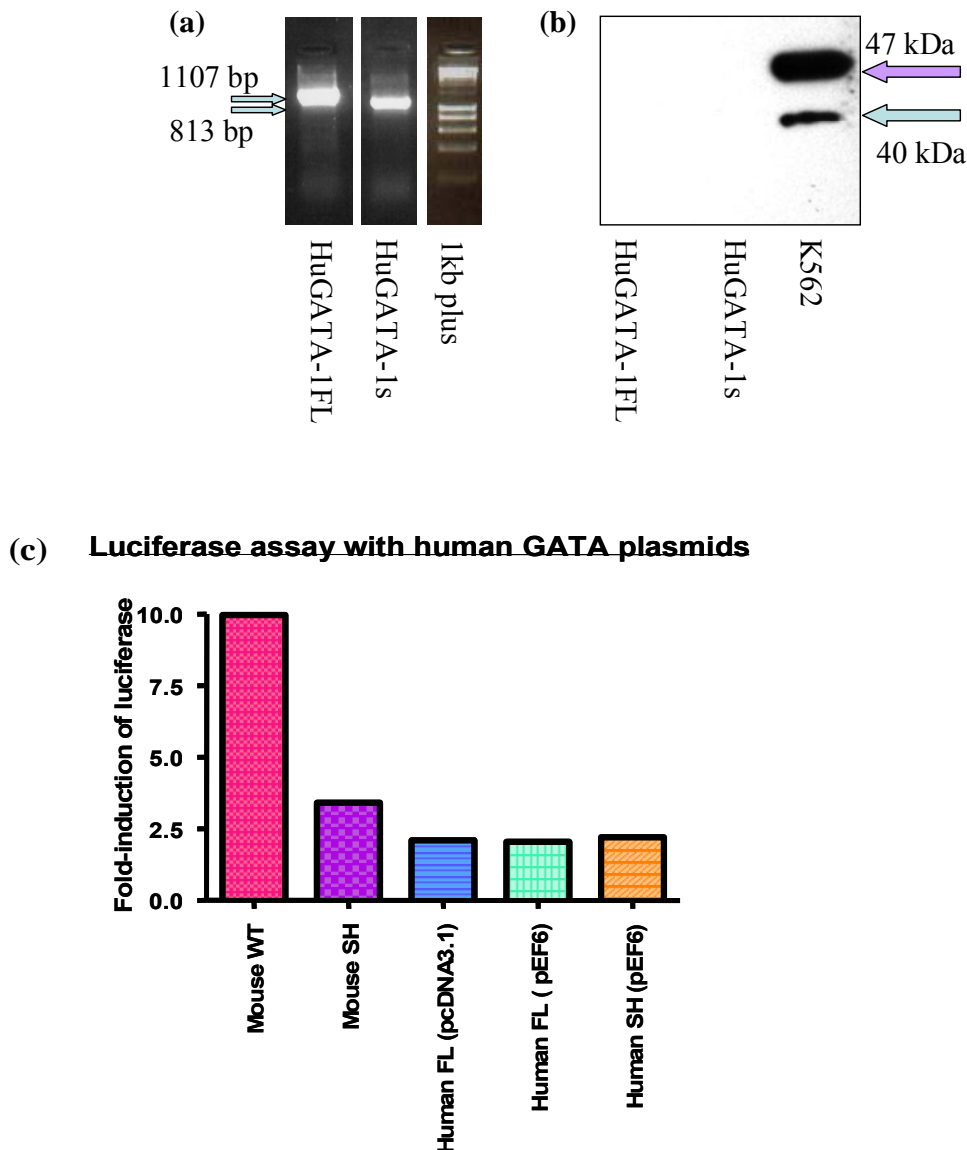
## **5.2 Construction of GATA-1 isoform specific expression vectors**

### **5.2.1 Human GATA-1 expression vectors**

The construction of human isoform specific expression vectors was facilitated by the presence of alternative splicing of GATA-1 mRNA in humans. Therefore, after RT-PCR of human cDNA from K562 cells using GATA-1 exon 1 and 6 primers (listed in table 2.5) bands corresponding to the GATA-1FL and GATA-1s isoforms could be directly excised, gel purified and TOPO cloned into a commercially available pEF6V5HisTOPO expression vector (Invitrogen) as described in section 2.12.1.1. This vector was chosen because it uses a human EF1 $\alpha$  promoter to drive gene expression which has been shown to be a strong promoter in haematopoietic cell lines (Goldman, *et al* 1996, Mizushima and Nagata 1990). Colonies were mini-prepped and analysed by restriction enzyme digestion for the presence of insert and its orientation. One putative GATA-1FL and 1 GATA-1s colony were sent for sequencing. Sequencing confirmed the presence of insert with correct orientation and confirmed 100% sequence identity with the published Human GATA-1 mRNA sequence (NM\_002049). This not only confirmed the correct sequence in the vectors but also revealed that the source of the template DNA - K562 cells (derived from a patient with chronic myeloid leukaemia) did not carry any mutations in the GATA-1 coding sequence.

The Human GATA-1 plasmids expressed GATA-1 at the mRNA level following transfection into HEK293 cells, as shown in fig 5.1a. However GATA-1 protein expression was not detectable by western blotting, despite multiple attempts (fig 5.1b). In addition, a luciferase reporter assay (described in detail in section 5.2.4.3) showed appropriate transactivation with the GATA-1s construct (2-3 fold induction of luciferase expression),

but failed to show enhanced transactivation with the full-length construct (fig 5.1c) despite adequate transactivation by murine GATA-1FL (described in 5.2.4.3). A number of attempts were made at recloning the GATA-1FL vector including subcloning it into an alternative expression vector pcDNA3.1V5HisTOPO (Invitrogen), adding a Kozak sequence (Kozak 1986b) to the FL GATA-1 upstream of the first ATG and using alternative template cDNA. However GATA-1 luciferase reporter assays continued to show only low level transactivation by the products from these plasmids (Fig 5.1c).



**Fig 5.1 Assessment of Human GATA-1 expression plasmids**

(a) The human GATA-1 plasmids produce GATA-1 mRNA as shown by RT-PCR using exon 1 and 6 primers which produce a 1107bp product with GATA-1FL and an 813bp product with GATA-1s (arrowed) (b) western blotting with the M20 antibody does not reveal a GATA-1 protein band despite GATA-1FL(47kDa) and GATA-1s (40kDa) bands in the positive control (K562) (c) a luciferase assay shows only modest transactivation of the reporter plasmid with equivalent levels in the human GATA-1FL and GATA-1s pEF6 plasmids (lanes 4 and 5). The GATA-1FL construct was recloned into a pcDNA3.1 vector but with no improvement in transactivation efficiency (lane 3). Levels of transactivation were equivalent to that seen with murine GATA-1s (lane 2) but well below murine GATA-1WT (lane 1).

It was decided not to pursue these vectors further. Instead murine GATA-1 isoform expressing plasmids were used for in vitro cellular differentiation studies in the human cell lines K562 and Meg-01. Although this species difference might theoretically interfere with interactions with GATA-1 cofactors and other transcriptional machinery in the cell, GATA-1 is known to be highly conserved between humans and mice. There is published work showing that expression of a murine GATA-1FL construct in a human haematopoietic cell line can produce terminal erythroid differentiation (Xu, *et al* 2003). The use of murine vectors does have potential advantages, particularly the ability to separately track exogenous vector derived GATA-1 (murine) from endogenous GATA-1 (human), and potentially the ability to use siRNA technology to knock-down human GATA-1 expression whilst leaving GATA-1 isoform expression from the murine transgene intact (as discussed in section 5.5.6 below).

### **5.2.2 Murine GATA-1 expression vectors**

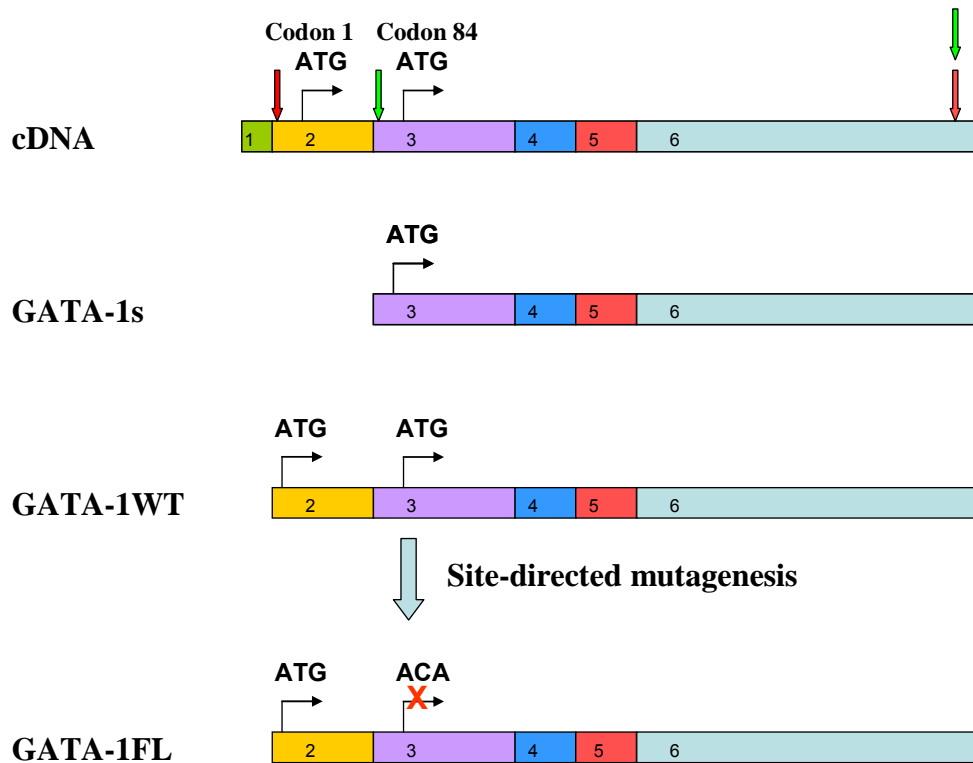
Since both GATA-1s and GATA-1FL can be transcribed from a single mRNA in mice it was necessary to modify this transcript to produce exclusive GATA-1FL expression. Fig 5.2 illustrates the design of these isoform specific transcripts. GATA-1s was produced by amplifying up murine GATA-1 cDNA starting just upstream of the exon 3 start codon (position 84), and therefore only allowing expression of the short transcript. In order to produce a vector capable of exclusive GATA-1FL expression an exon 2 forward primer was used, it was necessary to mutate the second start codon and therefore abolish initiation of translation at codon 84. The reverse primers annealed downstream of the exon 6 stop codon. All sequences are listed in tables 2.5 and 2.6.

### **5.2.3 Construction of pcDNAGATA-1FL, pcDNAGATA-1s and pcDNAGATA-1WT**

An N-terminal FLAG tagged murine GATA-1 cDNA (BFT4GATA-1) (kind gift from Prof Tariq Enver, Weatherall Institute of Molecular Medicine, Oxford) was used as the template DNA. The FLAG tag, a short hydrophilic 8-amino acid peptide (Asp-Tyr-Lys-Asp-Asp-Asp-Lys), used to efficiently detect protein expression by the use of high affinity anti-FLAG antibodies, was excluded from the final product as its N-terminal location could potentially interfere with the functional capabilities of this region of the molecule. Primers to amplify up the FL and short constructs are listed in table 2.5, chapter 2, and illustrated in Fig 5.2. PCR products were run on a 1% agarose gel, bands were excised and gel purified



and then TOPO cloned into the pCDNA3.1V5HisTOPO expression vector (Invitrogen) (fig 5.3a). Following restriction enzyme digestion to check for the presence of insert in the correct orientation, 2 promising clones (1 GATA-1WT (wild-type i.e. containing both start codons), 1 GATA-1s) were sent for sequencing (Technology services, Beatson Institute, Glasgow). 100% sequence identity with the published GATA-1 mRNA sequence (NM-008089) was confirmed for both clones.



**Fig 5.2 Creation of GATA-1 isoform specific products**

Arrows represent the position of GATA-1WT/FL primers (red arrows) and GATA-1s primers (green arrows). The GATA-1WT transcript underwent site-directed mutagenesis of the second ATG codon (red cross, ATG to ACA) to produce an exclusive GATA-1FL transcript.

The GATA-1WT clone was then used for site-directed mutagenesis of the second ATG start codon to produce the GATA-1FL vector. This involved a two base pair change in codon 84 from ATG (methionine) to ACA (threonine). This amino-acid change was suggested to be optimal by commercial site-directed mutagenesis software ([www.labtools.stratagene.com](http://www.labtools.stratagene.com)). Forward and reverse primer sequences are listed in Table 2.6. Site-directed mutagenesis was performed using a QuikChangeII commercial kit, as described in chapter 2 section 2.12.1.3. Successful introduction of the mutation was again confirmed by sequencing.

These vectors express the aminoglycoside phosphotransferase gene (neomycin resistance gene) under the control of the viral SV40 promoter which enables selection of stable transfectants by treatment of cells with the aminoglycoside antibiotic G418 (Geneticin).

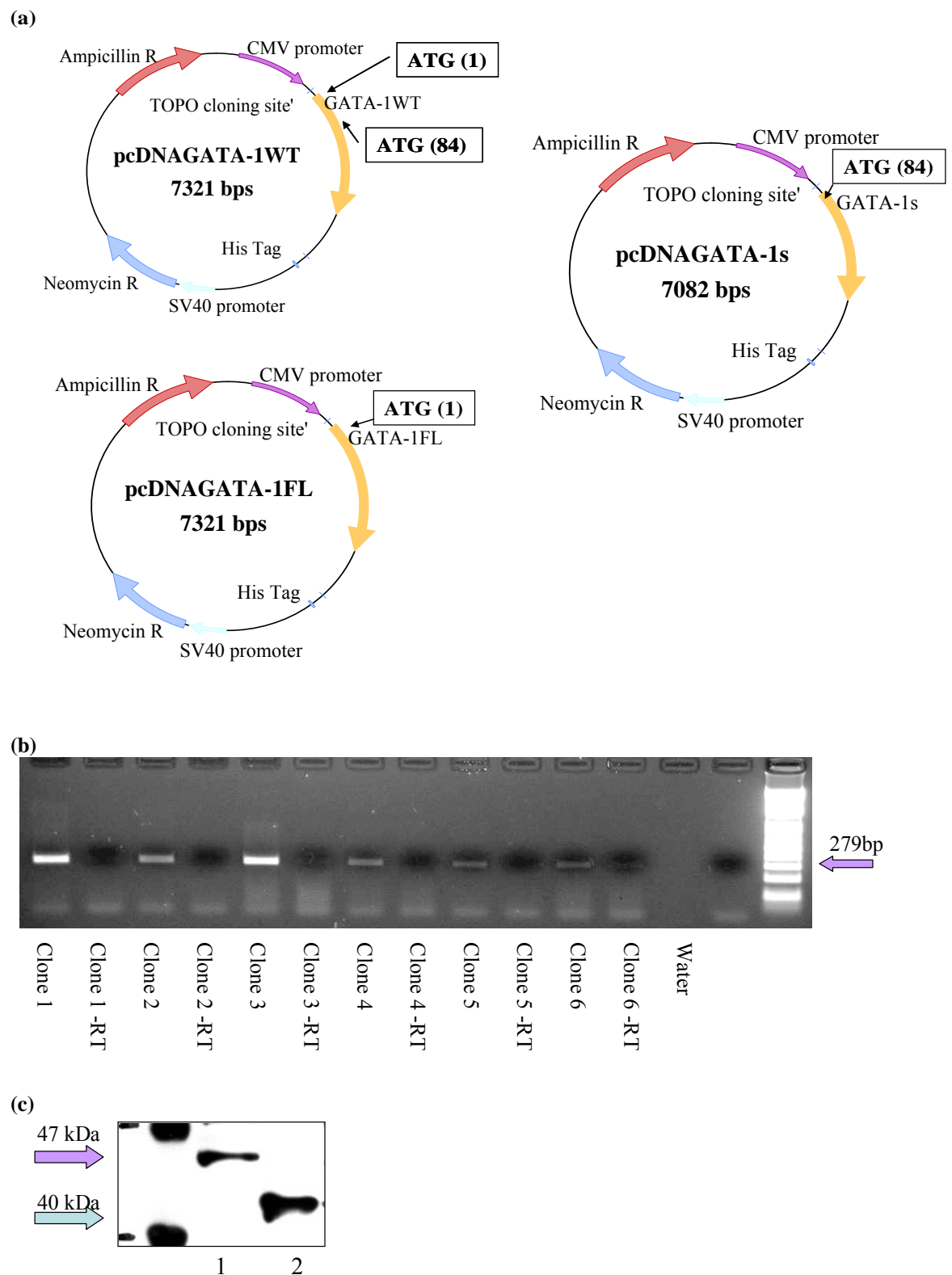
### **5.2.4 Validation of *pcDNAGATA-1* expression vectors**

#### **5.2.4.1 Expression at the mRNA level**

These vectors were then tested for GATA-1 mRNA expression by transient transfection into a cell line that does not normally express GATA-1 (HEK 293). Cells were seeded at a density of  $4 \times 10^5$ /ml in 10 cm plates the day prior to transfection. Transfection was carried out using 2 µg of plasmid DNA and the commercial transfection reagent Effectene (Qiagen) as described in section 2.12.3.3, chapter 2. Cells were harvested after 48 hours and RNA extracted using Trizol reagent (as described in section 2.8.1.1, chapter 2). Following DNase treatment (section 2.8.2), cDNA was made from the isolated RNA using a commercial kit (Superscript II, Invitrogen) as described in section 2.8.4, control reactions, excluding the reverse transcriptase (minus RT) were run for all samples. 2.5 µl of template DNA was used in a standard RT-PCR reaction using GATA-1 forward and reverse primers (table 2.4) which produce a product 279bp in size with amplification of cDNA (1937bp for genomic DNA). Results of this PCR are shown in fig 5.3b and confirm good levels of GATA-1 expression from all constructs with no evidence of expression in the minus RT controls (confirming results are not due to contaminating plasmid DNA).

#### **5.2.4.2 Expression at the protein level**

Having confirmed expression at the mRNA level, expression at the protein level was assessed. Fig 5.3c shows a western blot, using M20 antibody, of nuclear extracts made 48 hours after a second transient transfection experiment using HEK293 cells (this time transfected using a nucleoporation protocol described in section 2.12.3.4). This experiment confirmed exclusive GATA-1FL expression (47kDa upper band) from the pcDNAGATA-1FL plasmid and exclusive GATA-1s expression (40kDa lower band) from the pcDNAGATA-1s plasmid.



**Fig 5.3 Construction and expression from pcDNA vectors**

(a) Vector maps for pcDNAGATA-1WT, pcDNAGATA-1s and pcDNAGATA-1FL. Positions of start codons (ATG) are shown in boxes with codon positions (1 for FL and 84 for short) in brackets. (b) RT-PCR confirms expression of GATA-1 at the mRNA level, the primers produce a 279bp GATA-1 product (purple arrow), and lack of amplification from the vector is shown by absence of bands in the minus RT control lanes. (c) Western blotting confirms isoform specific GATA-1 protein production from the GATA-1FL plasmid, lane 1 (47kDa GATA-1FL protein detected, purple arrow) and the GATA-1s plasmid, lane 2 (40kDa GATA-1s protein, blue arrow) after transient transfection in HEK293 cells.

### 5.2.4.3 Production of functional protein

Finally, it was necessary to demonstrate that the proteins produced by the vector were functional and that the single amino-acid change at position 84 in the GATA-1FL expression plasmid did not alter GATA-1FL activity. To address this, a GATA-1 reporter plasmid (pGL3hGATA-1) was used (kind gift from Dr Paresh Vyas, Weatherall Institute of Molecular Medicine, Oxford). This plasmid was constructed with a number of GATA-1 binding sites upstream of a minimal promoter driving a Firefly luciferase reporter gene. Binding of GATA-1 to these upstream sequences should transactivate reporter gene expression. The Firefly luciferase gene takes part in a bioluminescent reaction catalysing the conversion of luciferin to oxyluciferin plus light. Light emission is proportional to the level of gene expression. Therefore, the level of luciferase gene expression can be assayed on a luminometer by measuring light production on addition of a commercial luciferin containing substrate as described in section 2.13.1, chapter 2.

Before assessment of luciferase transactivation could be made a number of parameters needed optimisation. In particular it was necessary to have a method for normalising for transfection efficiency to ensure that differing levels of light production did not merely reflect different levels of reporter and test plasmids within the cells. This necessitated co-transfection of a third plasmid encoding a  $\beta$ -galactosidase gene (pSV $\beta$ -gal). Luminescence results were then normalised to  $\beta$ -galactosidase activity. A number of optimisation experiments were performed:

1. To assess the optimum quantity and ratio of the three co-transfected plasmids ensuring that both luciferase and  $\beta$ -galactosidase reading remained within the linear range of their respective assays. Firstly, NIH3T3 cells seeded on 10cm plates were transfected with varying amounts of  $\beta$ -galactosidase vector (pSV $\beta$ -gal) (from 100ng to 1 $\mu$ g) and 1 $\mu$ g of a pGL3 control luciferase vector (pGL3Luc) (with constitutive luciferase expression). 100ng of  $\beta$ -galactosidase vector produced inconsistent results but amounts between 200ng and 1 $\mu$ g were able to consistently normalise light intensity measurements (data not shown).  $\beta$ -galactosidase standard curves (produced by serial dilution of a commercially provided  $\beta$ -galactosidase standard) were run with all assays to ensure measured levels of absorbance were within the linear range of the assay as described in section 2.12.4.1. A typical standard curve is illustrated in Fig 5.4a. Using a similar approach, co-transfection of differing amounts of reporter plasmid (pGL3hGATA-1) or test plasmids (pcDNAGATA-1FL) showed that maximal induction of luminescence could be achieved by using 4 $\mu$ g of test plasmid and 2 $\mu$ g of reporter plasmid.

2. To optimise transfection reagent conditions. Using a 4:2:1 ratio of test plasmid: reporter plasmid:  $\beta$ -galactosidase plasmid, established above, the optimal transfection protocol was then assessed. Fugene 6 was used in ratios from 1:1, 3:1, 3:2 and 4:1. A ratio of 4:1 produced the best transfection efficiencies (data not shown). This was used for all subsequent experiments

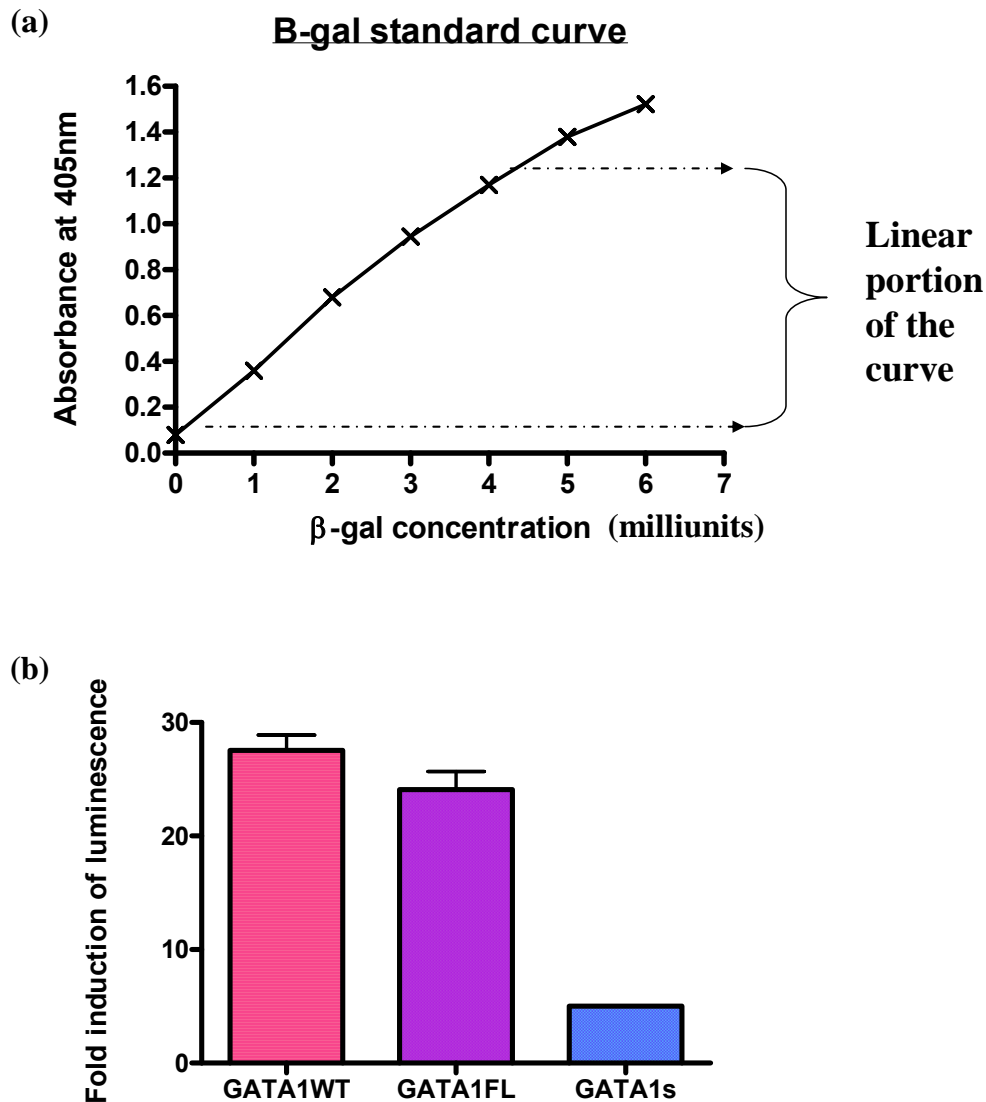
3. Finally, reporter assay performance was tested into two different cell lines– HEK293 and NIH3T3. Transfection into NIH3T3 cells produced good levels of luminescence (data shown below) but HEK293 cells were completely unable to sustain reporter gene activity (data not shown) despite good levels of protein expression by western blot (as shown in fig 5.3c above). This presumably reflects lack of essential co-factors for reporter gene transactivation in this cell type. Following this result all assays were carried out in NIH3T3 cells.

Following these optimisation experiments co-transfection into NIH3T3 cells of the three plasmids: pcDNAGATA-1FL, GATA-1s or wild-type (WT) (test plasmids), pGL3hGATA-1 (reporter plasmid) and pSV $\beta$ -gal (normalisation plasmid), was performed in triplicate. For each set of experiments untransfected cells were also assayed to control for background levels of luminescence and  $\beta$ -galactosidase activity, background expression from the pGL3hGATA-1 reporter plasmid was also assayed, and a positive control plasmid (pGL3Luc) was run to verify luminescence readings. Cells were harvested at 48-72 hours and cell lysates prepared using a commercial cell lysis buffer (reporter lysis buffer, Promega) as described in section 2.12.4.1. 30 $\mu$ l of cell lysate was used for the  $\beta$ -galactosidase assay (section 2.12.4.1) and 30 $\mu$ l for the luciferase assay (section 2.13.1). Results were initially corrected for background fluorescence (which was negligible) and background  $\beta$ -galactosidase activity, and then normalised to  $\beta$ -galactosidase activity as follows:

$$\frac{(\text{luminometer reading} - \text{background luminescence reading in untransfected cells})}{(\beta\text{-galactosidase absorbance reading} - \text{background absorbance in untransfected cells})}$$

Results were expressed as fold induction of luminescence (over levels obtained with the pGL3hGATA-1 reporter plasmid in the absence of test plasmid, arbitrarily set at 1.0) and are shown in Fig 5.4b. As can be seen in the figure both FL and wild-type GATA-1 vectors produce about a 20-fold induction of luminescence and the GATA-1s construct produces about a 5-fold induction over reporter alone. The difference between FL and wild-type levels is not statistically significant meaning that the mutation of the second ATG codon

did not produce a significant alteration in GATA-1 function as measured by this reporter assay. The difference between GATA-1s and GATA-1FL induction of reporter activity is consistent with published reports, and led to the original concept that the N-terminus was an important transactivation domain (Calligaris, *et al* 1995, Martin and Orkin 1990, Wechsler, *et al* 2002).



**Fig 5.4 Luciferase assay of GATA-1 transactivation potential**

(a) Representative  $\beta$ -galactosidase assay standard curve, brackets indicate range of values within the linear range of assay, any samples outwith this range were excluded (b) the GATA-1FL and GATA-1WT vectors produce high, and equivalent, levels of transactivation from the reporter plasmid, GATA-1s produces lower level transactivation (n=3)

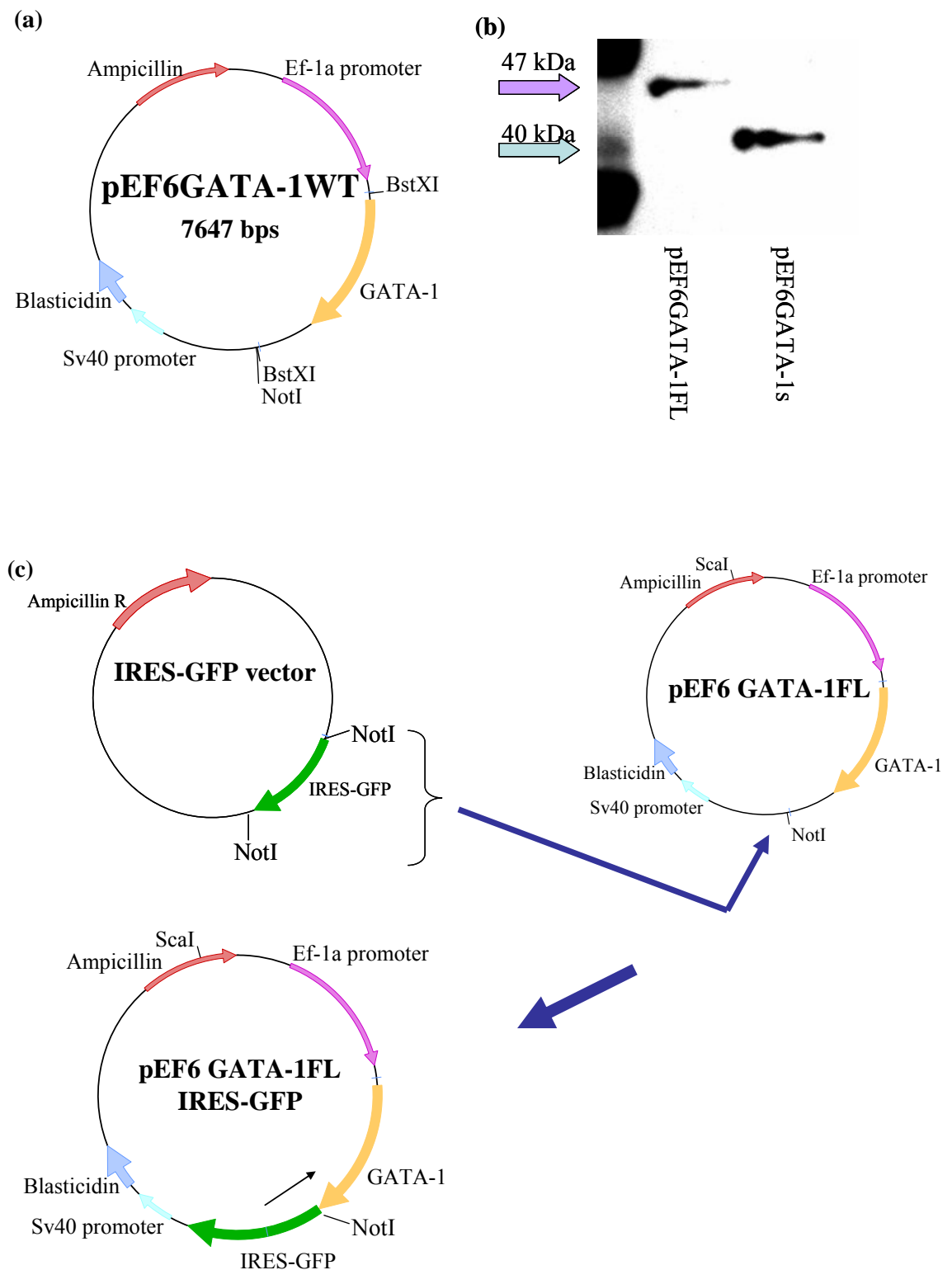
*These studies confirm that the murine GATA-1FL and GATA-1s expression vectors produce the expected functional protein products and that the introduction of a point mutation does not significantly alter GATA-1FL activity.*

## 5.2.5 Design of additional expression vectors

### 5.2.5.1 EFl $\alpha$ vectors

The pcDNA3.1 expression vector uses a CMV promoter to drive expression of the cloned insert in mammalian cells. This promoter is well established as suitable for expression of transgenes in a wide variety of cell lines, hence its choice for initial validation experiments and use in ES cell transfection (described in section 5.3 below). However, there are a number of reports of very poor expression levels using this promoter to drive transgene expression in haematopoietic primary cells and cell lines (Rodriguez, *et al* 2002, Salmon, *et al* 2000). Much better levels of expression are generally seen when using the human Elongation Factor 1 $\alpha$  (EFl $\alpha$ ) promoter which shows strong activity in human and murine cell lines (Goldman, *et al* 1996, Mizushima and Nagata 1990, Woods 2001). Therefore, to facilitate assessment of GATA-1 isoform expression in human haematopoietic cell lines, the GATA-1FL and short isoform cDNA sequences were amplified by standard PCR (using original BFT4 forward and reverse primers listed in table 2.5) from the pcDNA3.1 vectors and TOPO cloned into the alternative pEF6V5HisTOPO expression vector (Invitrogen) shown in fig 5.5a. Colonies were picked and the presence of insert and its orientation were assessed by restriction enzyme digestion (data not shown). Clones producing the correct restriction enzyme digest pattern were re-sequenced (Agowa.de) and these confirmed that no mutations had been introduced during the PCR process. Expression was confirmed by western blotting of nuclear extracts from transient transfection in HEK293 cells, using the M20 antibody (Fig 5.5b).

As shown in fig 5.5a, the pEF6 $\alpha$  vector expresses a resistance gene for the nucleoside antibiotic Blasticidin (Invitrogen) rather than G418. This necessitated determination of the minimal concentration of this antibiotic needed to kill the untransfected host cell line prior to its use in stable transfectants. This procedure is described in 2.12.3.1. For K562 cells a Blasticidin concentration of 10 $\mu$ g/ml was required, for Meg-01 8 $\mu$ g/ml and for E14 ES cells 5 $\mu$ g/ml.



**Fig 5.5 Cloning into the pEF6 expression vector and addition of IRES-GFP**

(a) vector map of pEF6GATA-1WT, note the EF1a promoter and Blasticidin resistance gene and the position of a NotI site downstream of the TOPO cloning site (b) Western blotting confirms isoform specific GATA-1 protein production from the GATA-1FL plasmid (47kDa GATA-1FL protein detected, purple arrow) and the GATA-1s plasmid (40kDa GATA-1s protein, blue arrow) (c) Ligation of an IRES-GFP construct into the pEF6 vectors using NotI sticky ended ligation to produce pEF6GATA-1FL IRES-GFP and pEF6GATA-1s IRES-GFP (not shown)



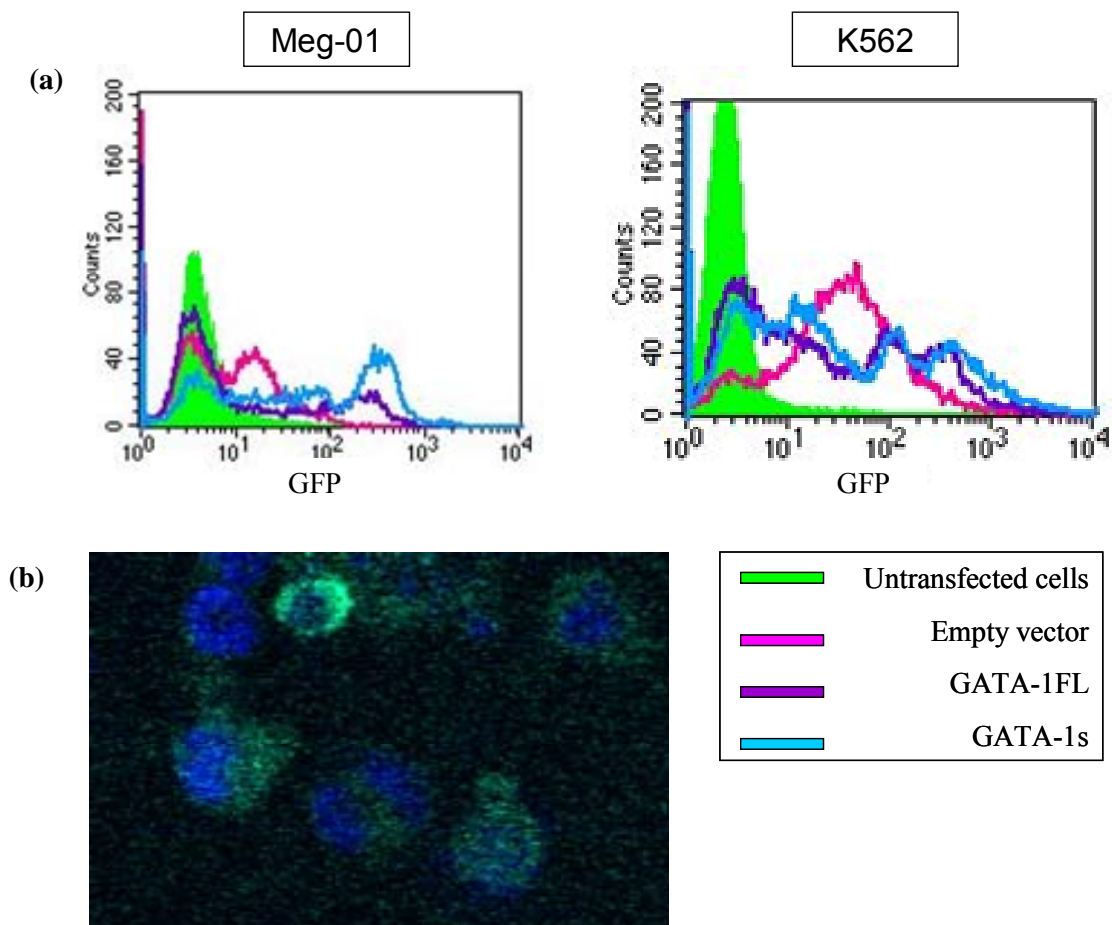
### 5.2.5.2 IRES-GFP vectors

The expression of some proteins in transfected cells can easily be assessed using western blotting, in situ hybridisation or flow cytometry. Unfortunately GATA-1 protein expression proved difficult to detect, requiring a nuclear extraction protocol and relatively large scale cell extracts (as described in section 3.1.2.1, chapter 3). To enable easier tracking and sorting of cells that had taken up the expression vector it was decided to add a green fluorescent protein (GFP) tag. GFP fluoresces green when exposed to blue light and this can be detected using the FL1 (FITC) laser channel in flow cytometers or a FITC filter set for fluorescent microscopy, allowing easy and rapid identification of transfected cells. Since the functions of the N- and C-termini were being evaluated it was decided not to directly fuse the GFP tag to the GATA-1 mRNA but instead introduce an internal ribosome entry site (IRES) sequence into the mRNA allowing formation of a bicistronic transcript capable of expressing both the intact GATA-1 protein and GFP (Mizuguchi, *et al* 2000, Mountford and Smith 1995). This sequence was subcloned from an IRES-GFP expression vector available in the laboratory (kind gift from Dr D Gilchrist, Division of Immunology, Infection and Inflammation, Glasgow University) whose IRES-GFP sequence was flanked by NotI sites. The pEF6 $\alpha$  vector also has a NotI site 3' to the TOPO cloning site allowing insertion of the IRES sequence at this point as shown in Fig 5.5c. NotI restriction enzyme digestion of both the EF1 $\alpha$  vectors (pEF6GATA-1FL and pEF6GATA-1s) and the IRES-GFP insert was performed. Products were gel-purified, the vector backbone was dephosphorylated (as described in section 2.12.1.2) to prevent religation and then the products were used in a standard ligation reaction (2.12.1.2). After colony picking, overnight culture and minipreps the presence of the insert was assessed using a NotI digest of plasmid DNA. 10/12 colonies for each isoform contained an insert. A KpnI digest was then performed to assess orientation and 3/10 GATA1FL colonies and 5/10 GATA-1s colonies had incorporated the insert in the correct orientation. One of each of these was 'maxi-prepped' using an endotoxin-free commercial kit for use in stable transfections; the vector could be linearised prior to stable transfection using a ScaI restriction enzyme site which cuts within the ampicillin resistance gene.

An empty vector control containing the IRES-GFP sequence but no upstream GATA-1 cDNA was also constructed as a negative control for transfection experiments. This was achieved by removal of the GATA-1s insert from pEF6GATA-1s using the BstXI restriction enzyme. This enzyme cuts at the sequence CCANNNNN/NTGG and therefore does not usually produce compatible sticky ends. To allow religation, the ends were

blunted (as described in 2.12.1.2.1) prior to use in a standard DNA ligation reaction. Following picking of colonies and screening with NcoI and PvuII digests 2/24 clones were found to have successfully excised the GATA-1s insert and religated. One of these was then digested with NotI and ligated with the NotI IRES-GFP insert in the same way as described for the pEF6GATA-1 vectors above. Again successful introduction of the IRES-GFP insert was confirmed with a KpnI digest which showed that 9/12 colonies had incorporated the insert in the correct orientation.

Expression of detectable GFP from these constructs was assessed following stable transfection into the haematopoietic cell lines K562 and Meg-01 (described in section 5.4.1 below). Flow cytometry was performed as described in section 2.13.2 - results are shown in fig 5.6a. Confocal microscopy images using nuclear counterstaining with DAPI (performed as described in section 2.13.6) are shown in 5.6b. Both these analyses produced similar findings i.e. the majority of cells expressed GFP but the level of expression varied considerably from cell to cell. This is discussed further in section 5.4.1.



**Fig 5.6 Expression of GFP in transfected cell lines**

(a) Representative FACS plots showing heterogeneous levels of GFP expression (b) confocal microscopy confirms GFP expression (green signal) by transfected K562 cells, nuclei are counterstained with DAPI (blue signal).

*In summary, this section describes the construction and validation of GATA-1 isoform specific expression vectors. Although human GATA-1FL and GATA-1s transcripts were easily amplifiable using RT-PCR, and subsequent cloning into vectors appeared successful (as judged by GATA-1 isoform mRNA expression with no errors on sequencing), it was impossible to demonstrate production of appropriately functioning human GATA-1 protein. Therefore murine vectors were pursued instead. 3 sets of murine vectors were produced using the CMV promoter (pcDNA3.1 vectors) or the EF1 $\alpha$  promoter with or without an IRES-GFP tag (pEF6 and pEF6-IRESGFP respectively). These vectors produced functional GATA-1 protein and the introduction of a point mutation at the second ATG did not significantly affect GATA-1FL function compared to wild-type GATA-1. The utility of the IRES-GFP tag in identifying transfected cells was demonstrated and a negative control plasmid, containing all the vector elements including the IRES-GFP cassette but without GATA-1 (empty vector control), was constructed. The use of these vectors to assess GATA-1 isoform function is described in the following sections, starting with murine ES cells followed by use in human haematopoietic cell lines.*

### **5.3 Expression of GATA-1 isoforms in murine ES cells**

Having constructed and validated the pcDNA3.1GATA-1FL and GATA-1s expression vectors, as described in section 5.2.3 above, these vectors were then linearised (using the ScaI restriction enzyme) and stably transfected into murine ES cells using Eugene 6 transfection reagent (see section 2.12.3.2). Cells were placed in selective media (G418 175 $\mu$ g/ml) at 24 hours post-transfection and media was changed daily. After 12 days colonies were sufficiently large to allow picking. 24 single colonies were picked per construct, as described in section 2.14.3. After growth in 24-well plates, 6 healthy looking, undifferentiated, colonies per construct were transferred to 6-well plates for expansion. These colonies were expanded up to 75cm<sup>2</sup> flasks and then harvested. Cell pellets were divided into three aliquots, 1/3 being used for RNA extraction (stored in Trizol reagent), 1/3 for protein extraction and 1/3 were frozen down and stored in liquid nitrogen for later use.

#### **5.3.1 Detection of GATA-1 expression**

Following RNA extraction (as described in section 2.8.1.1) and reverse transcription into cDNA using Superscript II (section 2.8.4) expression of GATA-1 at the mRNA level was

assessed by PCR. As shown in fig 5.7a GATA-1 expression was detectable in all clones but not the minus RT controls, confirming expression from the vector (note undifferentiated ES cells, such as these, do not show detectable endogenous GATA-1 expression as shown in Chapter 3 fig 3.12a). As might be expected with picking of individual colonies rather than pooled clones, the level of GATA-1 expression appears quite variable by standard (non-quantitative) PCR.

Expression at the protein level was not seen despite good positive control bands (K562) and loading controls (data not shown). As described in section 5.2.4.2 above, these same vectors were known to produce GATA-1 protein in other cell lines (fig 5.3c). Several possibilities exist to explain this lack of detectable expression in ES cells: (i) There could be low level protein production below the limits of detection of Western blotting with these GATA-1 antibodies (ii) there could have been technical problems with the vector or these particular transfected clones (iii) there could be post-transcriptional control of GATA-1 mRNA in ES cells (possibly by microRNAs or other mechanisms), preventing translation into protein. Of note, haematopoietic precursor cells are known to express multiple lineage specific genes at the mRNA level but not necessarily translate these into protein until a decision has been made to terminally differentiate (Hu, *et al* 1997).

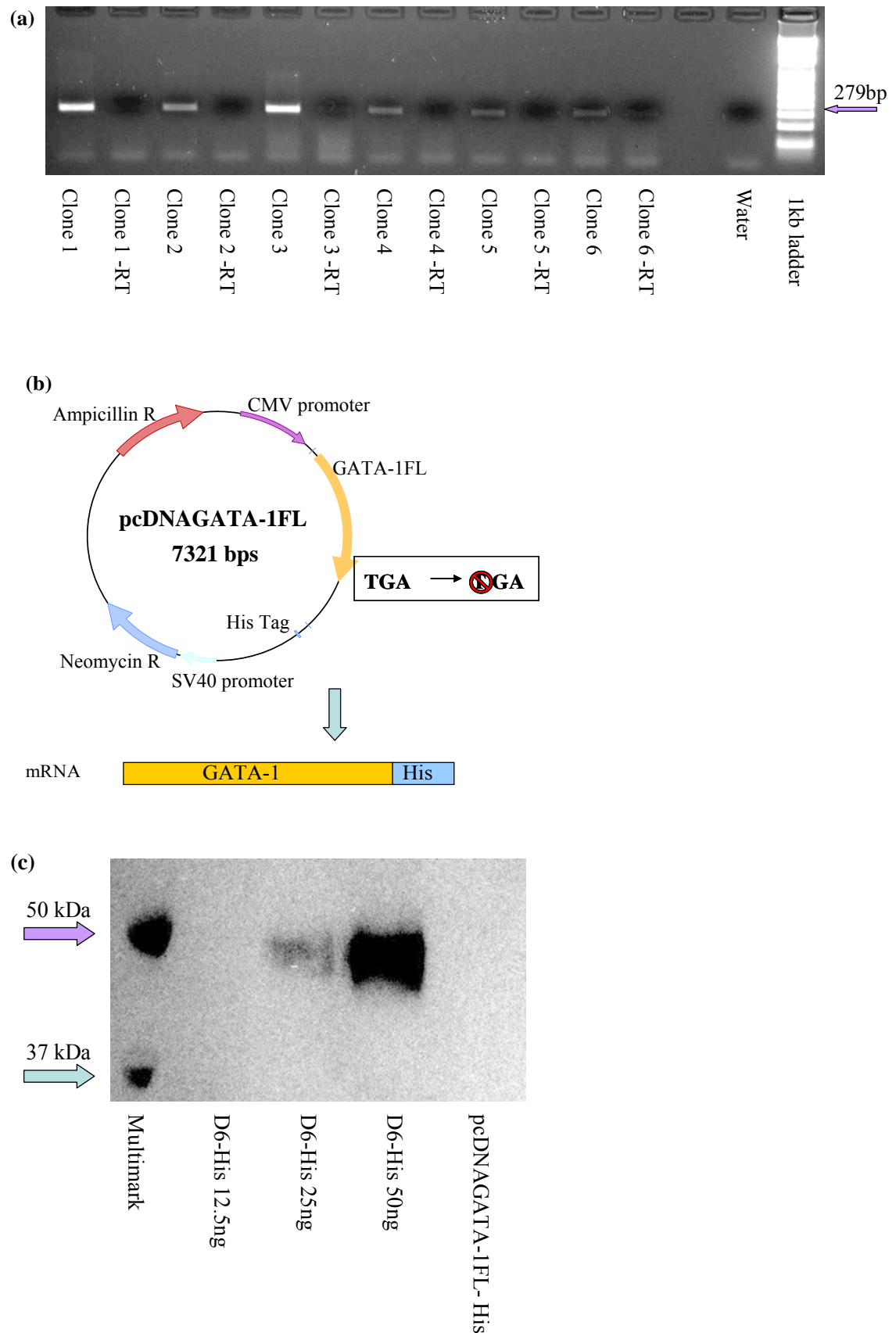
To test these possibilities several approaches were taken:

1. Nuclear extracts were made from a further batch of stable transfectants and the resulting antibody binding to western blot membranes was detected using an ultra sensitive chemoluminescent reaction (Supersignal West Femto). No expression was seen (data not shown).
2. To see if sensitivity could be improved by surrogate detection of GATA-1 expression using anti-His antibodies (table 2.13), site directed mutagenesis of the GATA-1 stop codon in the pcDNA3.1GATA-1FL plasmid was performed (as described in section 2.12.1.3, using primers listed in table 2.6). Primers were designed to delete the initial base of the TGA stop codon. This not only prevents termination of translation at this point but also produces a frameshift to allow run through translation to an in-frame His tag. This is illustrated in fig 5.7b. Despite adequate detection of a control His tagged protein (D6-His, available in the lab), no GATA-1His band was seen when this pcDNA3.1GATA-1His plasmid was stably transfected into ES cells (fig 5.7c).

3. To test whether lack of protein expression in the stable transfectants was due to transcriptional silencing of the integrated plasmid DNA, transient transfections were carried out to see if higher levels of expression could be achieved, resulting in detectable protein expression. These transient transfections were performed using the pEF6GATA-1FLIRES-GFP and pEF6GATA-1sIRES-GFP plasmids (section 5.2.5.2 above) to allow detection of GFP as well as GATA-1 expression in the transfected cells. Good levels of GFP expression from these vectors had previously been demonstrated in K562 and Meg-01 cells as shown in fig 5.6a. No detectable GFP expression (by FACS) from these vectors could be seen in ES cells despite reasonable transfection efficiency as shown by a control transfection with a GFP expressing plasmid (pMaxGFP). Western blotting on nuclear extracts from these cells confirmed no GATA-1 expression at the protein level (data not shown).

As already described in Chapter 3, section 3.2.7 there is evidence that endogenous GATA-1 production in ES cells is controlled at the post-transcriptional level. Fig 3.12 (chapter 3) shows good expression at the mRNA level (by RT-PCR) from day 3 of haematopoietic differentiation onwards, but no detectable protein production at day 6 (when mRNA levels appear highest).

Overall, these results suggest that exogenously introduced GATA-1 is transcribed into mRNA but protein production is either low-level and therefore undetectable by western blotting, or the mRNA is subject to post-transcriptional control in ES cells (or both). Despite this, there is strong evidence that these GATA-1 isoform expression vectors produce functional effects in ES cells as shown in the next section.



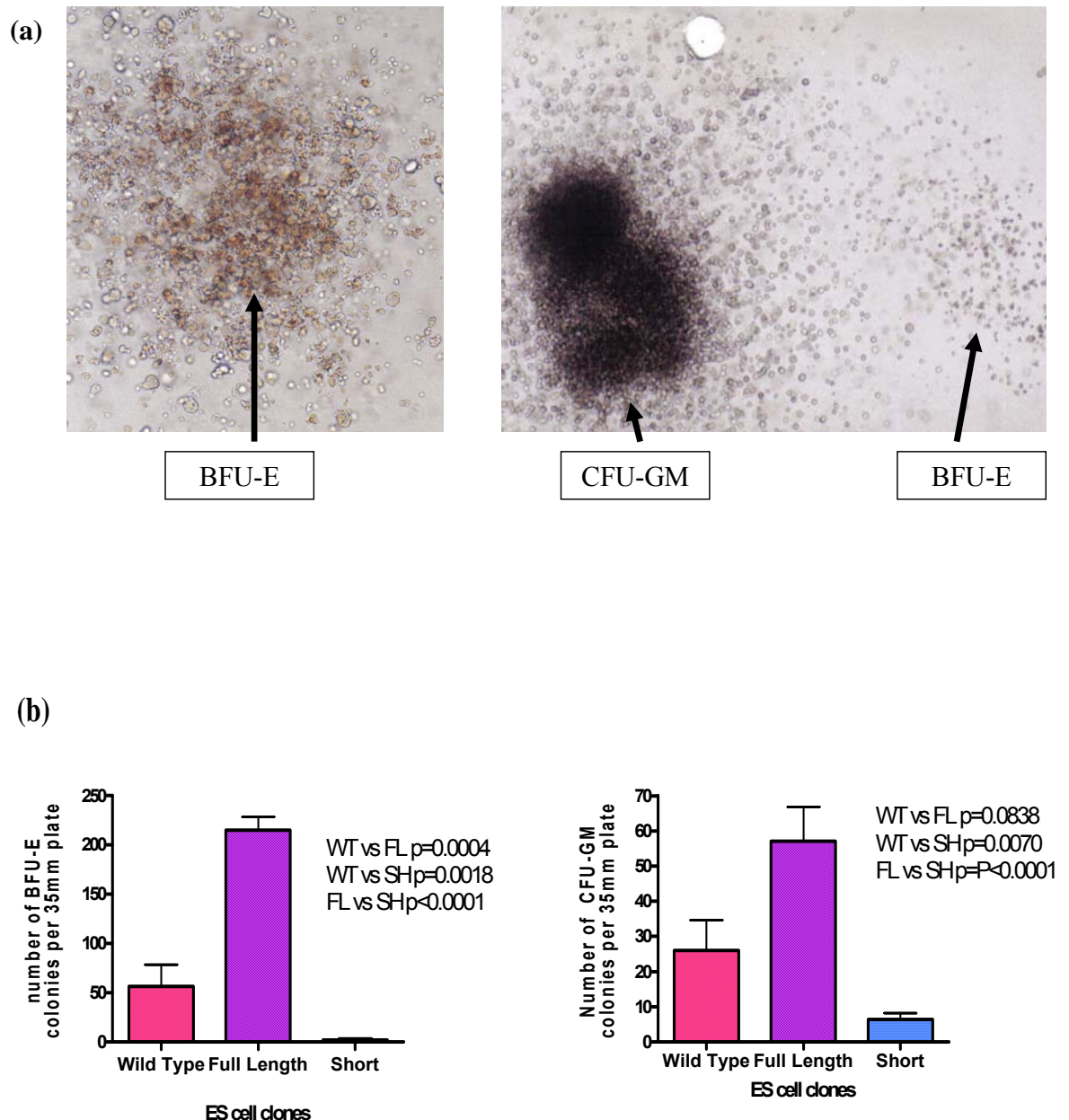
**Fig 5.7 Detection of transgene expression in murine ES cells**

(a) detection of GATA-1mRNA expression in ES cells by RT-PCR using exon 4 and 5 primers. The arrow shows the expected size of the 279bp GATA-1 product (b) site-directed mutagenesis of the pcDNAGATA-1FL plasmid to produce an in-frame His tag (c) western blotting reveals no detectable His protein from stably transfected ES cells despite good detection of another His tagged protein (D6-His) using the same antibody.

### **5.3.2 Effects of GATA-1 isoform expression vectors on haematopoietic differentiation of ES cells**

Despite low or absent protein expression from the vectors in undifferentiated ES cells, it is possible that protein production is switched on during haematopoietic differentiation and therefore it was decided to proceed to in vitro haematopoietic differentiation assays (note these vectors have previously been validated and are capable of producing functional protein as described in section 5.2.4 above). In order to assess whether the two GATA-1 isoforms produced differential effects on haematopoietic colony formation three stably transfected ES cell clones per isoform were used in methylcellulose assays. These assays involve plating cells in semi-solid media (methylcellulose) in the presence of growth factors and cytokines (Erythropoietin, Interleukin 3 and insulin), and in the absence of LIF. Under these conditions ES cells will produce haematopoietic colonies, primarily erythroid (BFU-E) and granulocyte/macrophage (CFU-GM). These colonies can be recognised by characteristic morphology as shown in fig 5.8a. The assay was performed, in triplicate, as described in section 2.14.2.2 and colonies were counted after 12 days. Fig 5.8b shows the number of colonies produced from the GATA-1FL and GATA-1s stable transfectants compared to untransfected E14 cells. As can be seen from this figure, clones transfected with the GATA-1FL construct produce a statistically significant increase in erythroid colony formation over both GATA-1s and untransfected ES cells. In addition there is a statistically significant decrease in both erythroid and non-erythroid (mainly CFU-GM) colony formation in GATA-1s transfected clones compared to both GATA-1FL and untransfected ES cells. These quantitative differences were also supported by qualitative differences: GATA-1FL BFU-E colonies were large and very deep pink or red due to haemoglobin production, whilst GATA-1s colonies were small, dispersed and pale pink (data not shown).

*In summary, experiments looking at GATA-1 isoform expression in murine ES cells show good levels of mRNA production from the GATA-1 introduced transgene but undetectable protein expression. Despite this, these transgenes produced functional effects when stably transfected into ES cells. Haematopoietic colony forming assays showed a significant induction of erythroid colony formation in GATA-1FL expressing clones, and a significant reduction in both erythroid and non-erythroid colony formation in GATA-1s expressing clones. This suggests GATA-1FL expression can accelerate erythroid development in this assay but that GATA-1s expression may produce a dominant negative effect preventing normal haematopoietic colony formation.*



**Fig 5.8 Effects of GATA-1 isoform expression on in vitro differentiation of murine ES cells**

(a) shows typical BFU-E and CFU-GM colonies produced by in vitro differentiation of ES cells in methylcellulose. BFU-E colonies tends to consist of small dense cells which are relatively scattered, without a dense core, they may appear pink/red due to haemoglobin production. GFU-GM are larger colonies with a dense core, individual cells are larger than BFU-E cells (photos courtesy of StemCell Technologies) (b) results of BFU-E and CFU-GM colony counts 10-14 days after ES cell in vitro haematopoietic differentiation ( $n=3$ ). Wild-type refers to untransfected E14 cells, Full-length = GATA-1FL transgenic ES cells, Short = GATA-1s transgenic ES cells

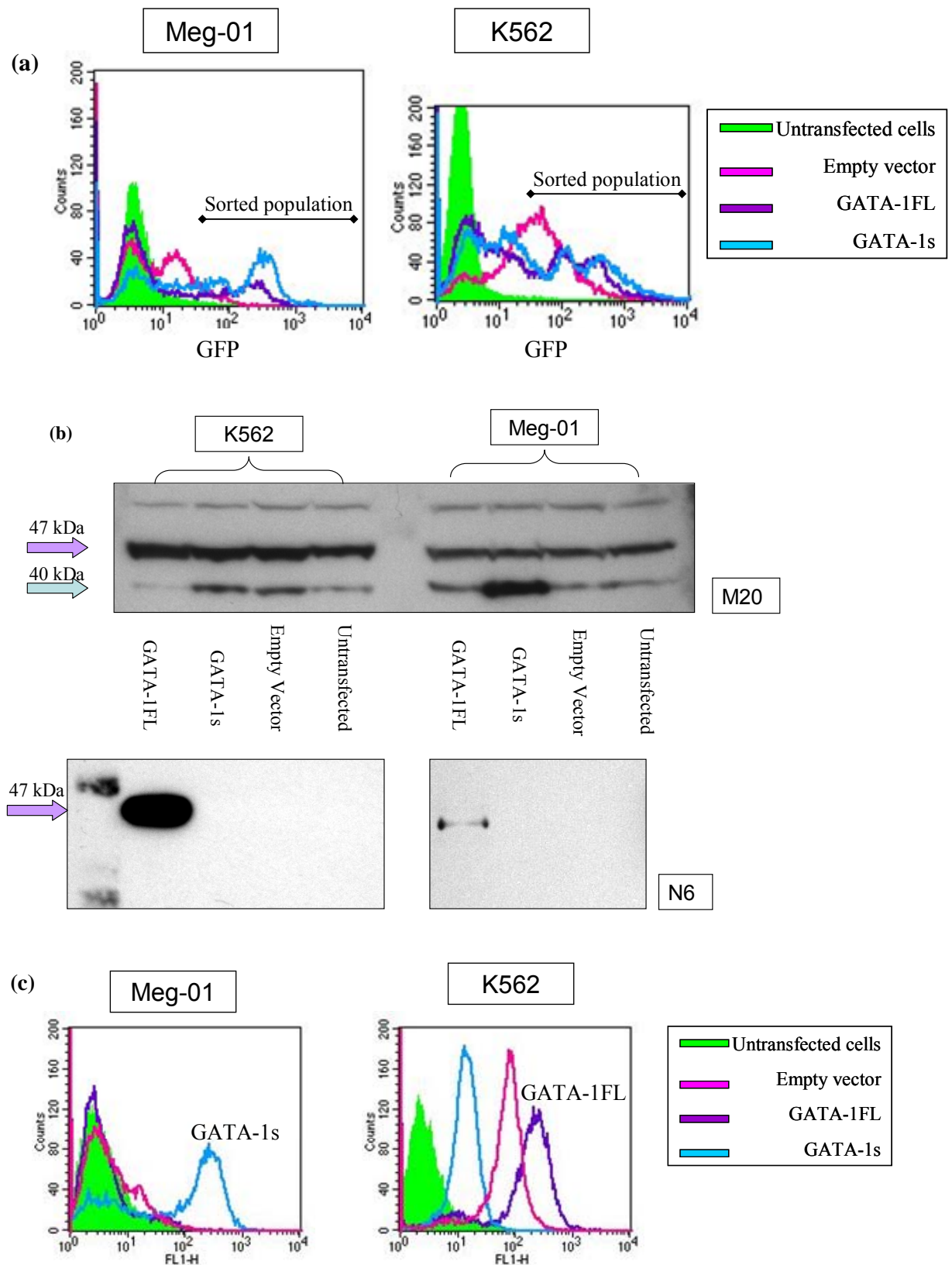


## **5.4 Expression of GATA-1 isoforms in K562 and Meg-01 cells – creation of transgenic lines and study of baseline characteristics**

The next step was to investigate the effects of GATA-1 isoform expression in human cell lines. Cell lines K562 and Meg-01 were both derived from patients with chronic myeloid leukaemia in blast crisis. Both cell lines are capable of differentiation along the megakaryocytic and erythroid lineages in vitro in response to inducing agents (Alitalo 1990, Lozzio and Lozzio 1979, Morle, *et al* 1992, Ogura, *et al* 1988). As previously mentioned, K562 is disomic for chromosome 21 whilst Meg-01 is trisomic. It should be noted that their derivation from CML blast crisis makes them potentially genetically unstable and that numerical differences in copies of chromosome 21 are not the only cytogenetic differences between the two cell lines. This makes attributing any differences in the biological behaviour of the two cell lines to differences in chromosome 21 number difficult. However, comparisons between these two cell lines provide the opportunity to uncover biological properties of the GATA-1s isoform only seen in the presence of trisomy 21, as suggested by the absence of leukaemic or pre-leukaemic phenotypes in mice (Li, *et al* 2005) and humans (Hollanda, *et al* 2006) carrying GATA-1s producing mutations without additional trisomy 21 (discussed in section 1.2.8). In addition, possible contributions of specific chromosome 21 genes to any observed differences could be formally tested by gene knockdown in Meg-01 cells or over-expression in K562 cells to see if this restores normal cell behaviour or reproduces the aberrant phenotype. Overall, these features make use of these two cell lines an attractive model for studying the effects of GATA-1 isoform expression on haematopoiesis.

### **5.4.1 Stable transfection**

K562 and Meg-01 cells were stably transfected with pEF6GATA-1FL IRES-GFP, pEF6GATA-1s IRES-GFP or pEF6empty IRES-GFP vectors (after linearization with *ScaI*) using nucleoporation (as described in section 2.12.3.4, Chapter 2). After 24 hours cells were placed in Blasticidin containing selection media and media was changed every 48 hours. After 1 week all cells on the control (untransfected) plate had died. An initial assessment of levels of expression from the transgene was made by FACS analysis looking for GFP expression from the IRES-GFP cassette (performed as described in section 2.13.2). Good levels of GFP expression were seen (as already shown in Fig 5.6a and reproduced here in Fig 5.9a), with a wide range of expression from low to high levels in



**Fig 5.9 Stable expression of GATA-1 isoforms in K562 and Meg-01 cells**

(a) FACS histogram of GFP expression soon after stable transfection showing a spectrum of GFP expression levels. Cells in the marked region were sorted (section 5.4.2) for later use (b) western blotting shows expression of FL and short isoforms in all cells with evidence of increased expression from the transgenic GATA-1 plasmid particularly evident for K562 GATA-1FL (as shown by N6 blot which only picks up murine GATA-1FL) and Meg-01 GATA-1s (M20 blot) (c) this western blot data is mirrored by spontaneous selection of cells for different levels of transgene expression on continuous culture. After 6 weeks in selection Meg-01 cells show high level GATA-1s expression whilst K562 show high level GATA-1FL.

individual cells. Western blotting confirms expression at the protein level as shown in Fig 5.9b – transgenic murine GATA-1FL expression can be confirmed since only the murine protein is detected by the N6 antibody. Expression of the GATA-1s transcript is inferred by the greater intensity of this band in the GATA-1s cell line compared to empty vector. GATA-1FL appears to be most strongly expressed in K562 (as seen on N6 lower blot) and GATA-1s appears to be most strongly expressed in Meg-01 cells (M20 upper blot). This is consistent with GFP expression data discussed below.

#### **5.4.2 Selection for high transgene expression**

To try and maximise biological effects of isoform over-expression it was decided to try and pick out high transgene expressers from amongst these pools of cells with variable levels of expression. High GFP expressers were cell sorted as described in section 2.12.5.  $2 \times 10^5$  K562 GATA-1FL, GATA-1s and empty vector containing cells with levels of GFP expression  $\geq 10^2$  in the FL1 (GFP) channel (population shown in fig 5.9a) were sorted and resuspended in 2ml culture medium in a 6 well plate. However with Meg-01 stable transfectants, despite initial sorting of  $2 \times 10^5$  cells, with no apparent technical difficulties, these cells did not survive in culture. Microscopic examination of these cells revealed prominent apoptosis. The sorting procedure was repeated with fresh stocks of Meg-01 transfectants with the same results. The empty vector control was affected by this process as much as the GATA-1s and GATA-1FL transfectants. This argues against cell death being due to high level expression of the GATA-1 transgene. It is most likely that Meg-01 cells, by virtue of their large size and megakaryocytic properties, may be intolerant of the sorting process. Another possibility is that high level GFP expression (rather than GATA-1) is toxic to Meg-01 cells; this has been reported in other cell lines (Liu, *et al* 1999, Misteli and Spector 1997).

#### **5.4.3 Spontaneous selection for isoform-specific transgene expression on continued cell culture**

Following these sorting experiments K562 high expressers and Meg-01 pools were maintained in selective media for several cell passages. Prior to their use in in vitro differentiation experiments (described in section 5.5 below), GFP expression was reassessed by FACS analysis. Interestingly, as shown in fig 5.9c, the cells had spontaneously segregated themselves according to the expressed transgene i.e. in Meg-01 cells there was no longer a spectrum of GFP expression as shown in fig 5.9a, instead cells

carrying the GATA-1FL transgene had spontaneously selected for low levels of transgene expression, whilst cells carrying the GATA-1s transgene had spontaneously selected for high level expression. The empty vector showed intermediate to low level expression. The K562 cells showed opposite effects i.e. despite, in this case, the sorting of these populations for high level GFP expression, GATA-1FL transfectants had maintained high level expression but GATA-1s transgenic cells has down-regulated expression with empty vector remaining at intermediate levels. The down-regulation of introduced transgenes with continued cell culture (despite maintenance in selective medium) is a well recognised phenomenon (Pikaart, *et al* 1998, Recillas-Targa 2006) and may explain the low-intermediate levels of the empty vector transgene. However the maintenance/selection for high levels or particularly low levels of expression suggests that the GATA-1 isoforms are having selective effects on cell growth or survival, which differ in the two cell lines. This may be due to effects on cell proliferation or prevention or induction of terminal differentiation i.e.

1. High GATA-1s expression appears to have a selective advantage in Meg-01 cells (trisomy 21) but not K562 (disomy 21). It may be that GATA-1s is able to promote survival/proliferation of immature megakaryocytic precursors in Meg-01 cells and prevent terminal differentiation. Since this same phenomenon is not seen in K562 cells, it is tempting to speculate that the presence of trisomy 21 unmask this potential, but this is unproven.

2. High GATA-1FL expression appears to be advantageous in K562 cells whereas GATA-1s expression appears to cause a growth or survival disadvantage. K562 cells are less committed than Meg-01 cells and the expression of GATA-1FL in these cells may promote cell expansion and initial commitment to meg-erythroid lineages. By analogy with the ES cell methylcellulose assay (described in section 5.3.2) high levels of GATA-1FL could promote progenitor expansion whilst high levels of GATA-1s isoform expression could have a dominant negative effect. This would lead to selection for high level GATA-1FL and low level GATA-1s expressers when K562 cells were maintained in culture.

#### **5.4.4 Methods of evaluation of baseline characteristics**

In order to explore these phenomena further and establish baseline characteristics of the transgenic cell lines, prior to their use in in vitro differentiation experiments (described in section 5.5), a number of experiments were performed:

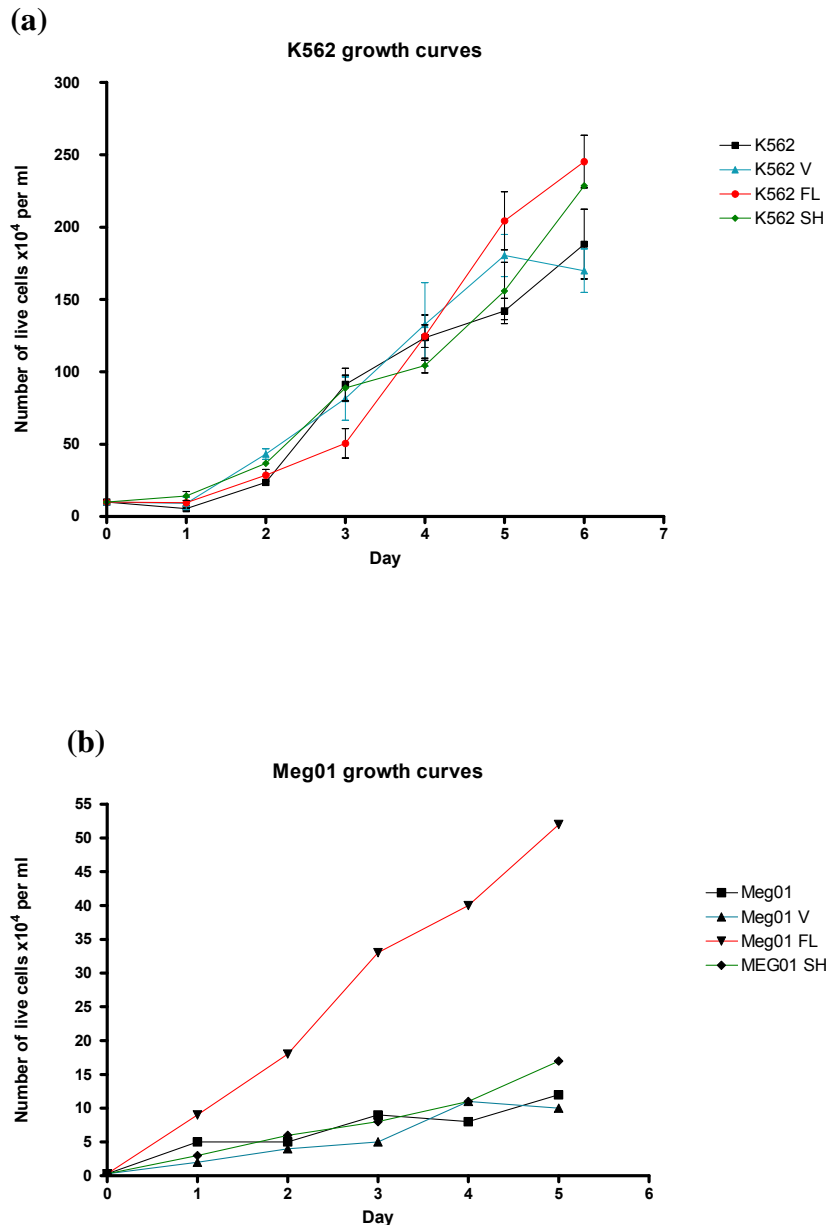
1. Measurement of cell growth rates in the presence of the different GATA-1 isoforms
2. Examination of maturation status of the transfected cells under steady state conditions using:
  - (i) cytopins to assess cell morphology and multinuclearity.
  - (ii) measurement of the DNA ploidy of cells in steady-state culture
  - (iii) analysis of erythroid and megakaryocytic cell surface marker expression
3. Analysis of gene expression profiles of key GATA-1 interacting genes under steady-state culture conditions.

The results of these analyses are presented below:

#### **5.4.5 Cell growth rates**

To assess whether the presence of the transgene had any affect on cell growth rates a simple cell growth assay was performed. Cells were harvested and seeded at  $1 \times 10^5$  cells per ml in 6 well plates for K562 cells and  $1 \times 10^4$  cells per ml for Meg-01, assigning one well per construct per day. Each day the corresponding wells were harvested and the number of live cells was estimated using a modified Neubauer counting chamber with Trypan blue exclusion of dead cells (as described in section 2.5.2). Particular care was taken to scrape the wells prior to harvesting to resuspend any loosely adherent cells. Wells for days 4 onwards were re-fed on day 3 to prevent exhaustion of nutrients in the culture medium, cell counts were adjusted to account for this larger media volume from day 4 onwards. Results are shown in Fig 5.10a and b for K562 and Meg-01 respectively. In addition to the GATA-1FL, GATA-1s and empty vector constructs, untransfected K562 and Meg-01 cells were included in this analysis. This was done to ensure that the expression of GFP (from the empty vector) did not interfere with cell growth characteristics. As can be seen in Fig 5.10 the growth curves for the untransfected cells and the vector only controls were very similar with no statistically significant difference between the two cell populations in K562 cells (using 2 way ANOVA analysis  $p=0.599$ ). The other two constructs – GATA-1FL and GATA-1s- both provided a modest but significant growth advantage by day 6 compared to empty vector (using a two-tailed students t-test  $p=0.033$  for GATA-1FL vs. empty vector control and  $p=0.018$  for GATA-

1s vs. empty vector) but did not produce a statistically significant difference when compared to one another (GATA-1s vs. GATA-1FL  $p=0.40$ ). In contrast GATA-1FL transgene expression produced a clear increase in cell growth rates in Meg-01 cells compared to both GATA-1s and empty vector/untransfected cells. GATA-1s, contrary to predictions made from GFP expression levels (section 5.4.3) appeared not to produce a growth advantage in Meg-01 cells. There was no difference in the proportion of dead cells (as measured by Trypan blue staining) seen with any of the constructs.



**Fig 5.10 K562 and Meg-01 growth curves**

Growth curves for (a) K562 ( $n=3$ ) and (b) Meg-01 ( $n=1$ ) cell lines either untransfected (wild-type) or containing empty vector (V), GATA-1FL (FL) or GATA-1s (SH) transgenes.

*Therefore it appears that expression of the GATA-1FL transgene gives Meg-01 cells a proliferative advantage in cell culture, whilst GATA-1s behaves in a similar way to empty vector controls. In the K562 cell line both the GATA-1FL and GATA-1s transgenes produce a moderate growth advantage over the empty vector cell line. These data support a role for GATA-1 in promoting cell survival and proliferation of meg-erythroid precursors. They also suggest that GATA-1FL may be more dominant than GATA-1s in this regard in a cell line specific manner. They do not, however, shed light on the mechanism of selection for high GFP expression in Meg-01 GATA-1s expressing cells.*

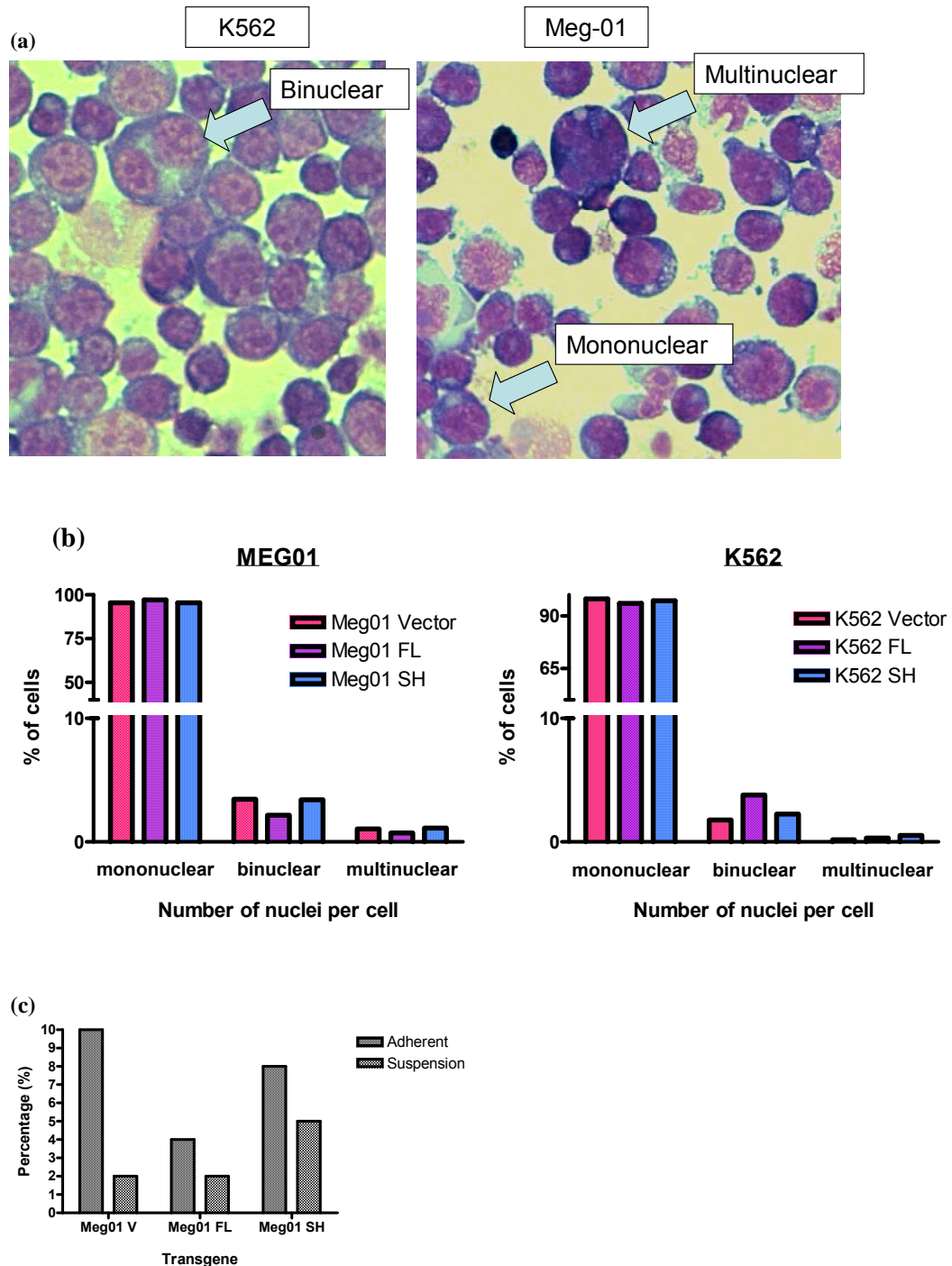
#### **5.4.6 Maturation status of cells**

Besides affecting cell proliferation the GATA-1 transgenes may have an affect on maturation of the Meg-01 and K562 cell lines. To assess whether maturation status varied between steady state (i.e. without induced differentiation) cells a number of different assays were used as detailed in section 5.4.4. Results are presented below:

##### **5.4.6.1 Evaluation of cell morphology**

Cells were harvested, counted and washed in PBS before being attached to glass slides using a cytopsin technique as described in section 2.13.5. Slides were air-dried and then stained with May-Grunwald-Giemsa stain. All slides were viewed using a conventional light microscope equipped with a cell imaging system (Axiostar plus, Carl Zeiss). Representative cell morphology is shown in Fig 5.11a. Megakaryocytic maturation is associated with endomitosis leading to multinuclearity. To quantitate this, the numbers of mononucleate, binucleate and multinucleate cells were counted in 10 high powered fields (minimum 200 cells). Representative examples are labelled in Fig 5.11a. As expected Meg-01 cells had slightly higher numbers of multinuclear cells than K562, but there was no overall difference between the different transgenes in the proportion of mononuclear, binuclear or multinuclear cells as shown in Fig 5.11b. Since Meg-01 cells are loosely adherent, it might be hypothesised that more mature cells (which are large and produce long processes, with occasional visible platelet budding) reside in the adherent portion and may be missed by cytopsin made from cell suspensions. To test this hypothesis, cytopsin were performed on the adherent proportion (isolated by removal of the supernatant, washing twice in PBS, adding 1ml of sterile PBS and then mechanically scraping the adherent cells before collecting by centrifugation). Cell counts performed on adherent cells did show an increased number of multinuclear cells compared to suspension controls

(average 8% of adherent cells were multinuclear versus 3% of suspension cells) as shown in Fig 5.11c. Both the GATA-1 transgenes (GATA-1FL and GATA-1s) seemed to be associated with lower numbers of multinuclear cells in the adherent fraction than the empty vector control, although the significance of this is unclear.



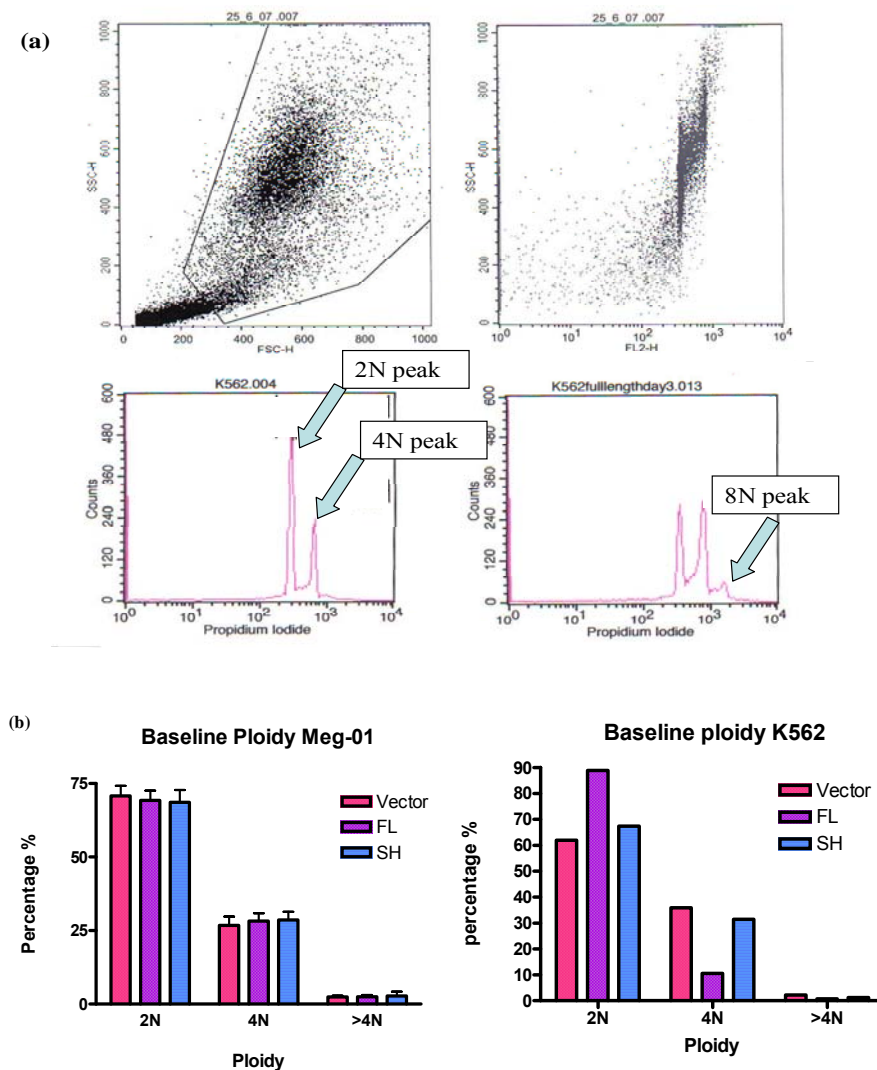
**Fig 5.11 Influence of transgene expression on cell morphology**

(a) cytopspins of K562 and Meg-01 stable transfectants, examples of mononuclear, binuclear and multinuclear cells are indicated by arrows (b) percentage of each cell type counted in 10 high power fields (minimum 200 cells), no significant differences between the transgenes were seen (n=3) (c) percentage of multinucleate cells in adherent and suspension populations (n=1). The adherent fraction is associated with increased numbers of multinucleate cells.



### 5.4.6.2 Measurement of DNA ploidy

Another method of assessing the degree of endomitosis (and therefore megakaryocytic maturation status) is to measure the DNA content of individual cells. Endomitosis involves an increase in DNA content with successive cycles of nuclear replication. This leads to an exponential increase in ploidy from 2N to 4N to 8N to 16N etc. Ploidy can be measured by flow cytometric evaluation of the intensity of cell staining with the fluorescent DNA intercalating agent propidium iodide. Cells were prepared and stained as described in section 2.13.2.1, Chapter 2. A cell line with known 2N DNA content (the Burkitt's lymphoma derived cell line – Raji, cultured as described in table 2.2) was included as a positive control. Results are shown in Fig 5.12 below:



**Fig 5.12 Ploidy analysis of K562 and Meg-01 stable transfectants**

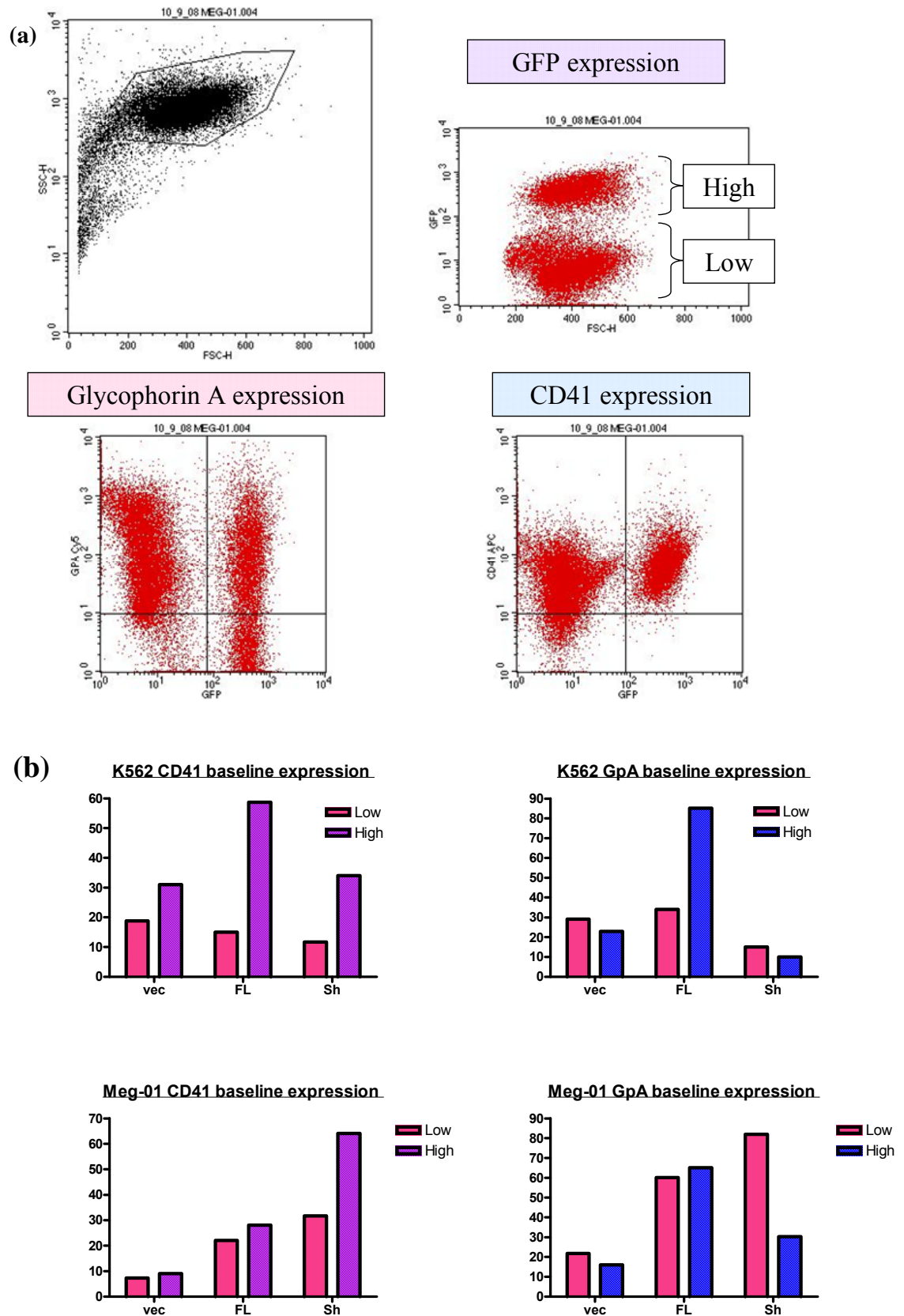
a) Representative FACS plot (scattergram and histogram) showing location of 2N, 4N and 8N peaks b) baseline DNA ploidy for Meg-01 (n=3 replicates, no significant differences seen between constructs) and K562 (n=1) cells, expressing the 3 transgenes (empty vector, GATA-1FL and GATA-1s)

As can be seen from this figure the three transgenes do not appear to have substantial differential effects on baseline ploidy, with the possible exception of reduced numbers of 4N cells in the K562 GATA-1FL cell line. This result differs from the cytopsin findings (which actually showed increased binucleate cells in K562 GATA-1FL) and its significance is unclear. In Meg-01 cells there were no statistical differences in ploidy class between the 3 transgenes.

#### **5.4.6.3 Analysis of cell surface marker expression**

As megakaryocytic cells mature they should upregulate essential platelet proteins such as the glycoprotein complex GpIIb/IIIa. Similarly, erythroid maturation is associated with increased expression of cell surface proteins such as Glycophorin A. Therefore expression of these proteins can be used as surrogate markers for cell maturation. Fluorescently labelled monoclonal antibodies were used to assess GpIIb/IIIa (antiCD41a-APC) and Glycophorin A (anti-CD235a-PE-Cy5) expression as described in chapter 2 section 2.13.2. These fluorophores were chosen to minimise overlap with transgenic expression of GFP (from the IRES-GFP sequence in the expression vectors). Despite this it was necessary to perform compensation prior to analysis to minimise interference between the different channels, this was done with the help of experienced FACS users (Dr A Fraser and Dr C Hansell, Division of Immunology, Infection and Inflammation, Glasgow University). Cells were initially gated on FSC and SSC to encompass a viable bulk cell population; they were then gated on GFP expression to allow correlation between the level of transgene expression and the measured cell surface marker. The levels of Glycophorin A and CD41 expression and number of positive cells were then assessed. Fig 5.13a shows representative FACS plots. Fig 5.13b presents these results in graphical form.

As can be seen from these plots both cell lines expressed detectable levels of both erythroid (GpA) and megakaryocytic (CD41) cell surface markers. The expression data are interesting. In K562 cells all three transgenes show similar levels of expression in the GFP low cells (as would be predicted, as transgene expression is assumed to be low in this population and they should therefore resemble the empty vector controls). In the GFP high cells the FL isoform produces marked upregulation of CD41 and GpA. The GATA-1s transgene is associated with less upregulation of CD41 and down-regulation of GpA in the GFP high group. In Meg-01 cells it is the GATA-1s isoform that is associated with the highest levels of CD41 both in the low and high GFP gates but again increases in GFP expression are associated with down-regulation of GpA. This suggests that the GATA-1FL transgene promotes expression of megakaryocytic and erythroid markers in both cell lines



**Fig 5.13 Expression of erythroid and megakaryocytic cell surface markers in Meg-01 and K562 transgenic cell lines**

(a) representative FACS plot for Meg-01 cells showing high and low GFP expressing populations and variation in GpA and CD41 expression (b) Baseline CD41 and GpA expression expressed as mean fluorescence intensity for the high GFP (high) and low GFP (low) gated populations (n=1)

(although most prominently in K562), consistent with its ability to drive differentiation of cells down either lineage. The GATA-1s isoform appears to favour expression of megakaryocytic (CD41) over erythroid markers (GpA) and appears to suppress GpA expression when present at a high level. This effect on GpA is particularly marked in the Meg-01 cell line (which carries trisomy 21).

*In summary, these 3 methods to assess maturation status of cells revealed little or no differences in baseline numbers of multinucleate cells or DNA ploidy between any of the transgenes, in either cell line. Cell surface marker expression suggested that the GATA-1FL isoform drove expression of both megakaryocytic and erythroid markers (CD41 and GpA respectively) in both cell lines. The GATA-1s isoform showed little CD41 upregulation in K562 but marked upregulation in Meg-01 especially in high GFP expressors. GATA-1s CD41 upregulation was associated with GpA down-regulation in both cell lines. This suggests that high GATA-1s expression down-regulates markers of erythroid commitment and enhances CD41 expression consistent with megakaryocytic commitment. These data are consistent with the findings in murine ES cells of accelerated erythroid and CFU-GM commitment in GATA-1FL transgenics and suppression of erythroid colony formation in GATA-1s transgenics. The lack of comparable differences in the cytospin and ploidy analysis suggests that this high CD41 expression with GATA-1s is not accompanied by other signs of megakaryocytic maturation. However it may also reflect the use of bulk cell populations, in contrast to the gating on high GFP in the cell surface marker studies.*

#### **5.4.7 Gene expression profiles of transfected cell lines**

To explore the consequences of transgenic GATA-1 isoform expression further it was decided to examine their effect on gene expression profiles in the two cell lines. As extensively discussed GATA-1 is a key haematopoietic transcription factor whose expression leads to alteration in the transcription of a very large number of genes either directly or via a cascade of intermediate factors. The genes whose transcription may be affected by GATA-1 can be classified into 3 broad groups:

1. Lineage restricted genes - the vast majority of megakaryocytic and erythroid specific genes possess GATA-1 binding sites within their regulatory regions and it is likely that binding of GATA-1 to these sites is necessary to institute lineage specific gene expression.

2. Genes involved in alternative cell fate decisions - in addition GATA-1 influences the transcription of master regulators of other lineages. This is consistent with its role, not only in specifying megakaryocytic and erythroid cell fate, but also switching off alternative lineage choices such as lymphoid and granulocytic differentiation.
3. Genes involved in proliferation and cell cycle control. The initial observations of GATA-1 knockout cells revealed its essential function in promoting cell survival and proliferation as well as allowing terminal differentiation.

The clinical phenotype of GATA-1s knockout mice, the human GATA-1s kindred and Down syndrome children with TAM/AMKL, suggest abnormalities in maturation, lineage choice and cell proliferation and therefore it is likely that the GATA-1s isoform shows differential effects on gene expression from all three of these categories. Initial data from microarray studies in Down syndrome associated disease (Bourquin, *et al* 2006, Ge, *et al* 2006) and the GATA-1 transgenic mice (Li, *et al* 2005) has suggested a number of candidate genes that may be differentially regulated by the GATA-1s isoform (listed in table 5.1 below). Because of this, it was decided not to perform a further microarray study, but rather to concentrate on a smaller number of potentially informative genes to examine their response to GATA-1 isoform expression in more depth using quantitative PCR. To facilitate this micro-fluidic plate technology was utilised in the form of Taqman low density array (TLDA) plates. These plates consist of 384 wells preloaded with standardised primers and probes. These wells are connected via micro-fluidic channels allowing centrifugal loading of template and PCR master mix into multiple wells via one sample port. This technology ensures high reproducibility between plates, since variability in inter-well pipetting, primer and probe mixing, and sample loading is virtually eliminated. The format chosen for this set of experiments was the assay of 15 genes plus one endogenous control gene (18SrRNA) with each assay being performed in triplicate. Therefore each plate is capable of analysing 8 biological samples ( $8 \text{ (samples)} \times 16 \text{ (genes)} \times 3 \text{ (replicates)} = 384$ ). The choice of genes was based on promising published data or key roles in lineage determination likely to show differential regulation by the two isoforms. These genes are listed in Table 5.1 along with the reasons for their inclusion.

Gene expression in each biological sample was assayed in triplicate and 3 biological replicates for each cell line/time point were run on the plate allowing statistical comparison between different biological samples where appropriate. Samples were placed in RLT buffer (RNeasy kit, described in section 2.8.1.2) and frozen at  $-80^{\circ}\text{C}$  until ready for full RNA extraction. In order to minimise inter-assay variation due to reagent/kit differences,

Gene	Class	Existing microarray data	Reason for inclusion
Erg3	Chromosome 21 gene Meg differentiation	None	Expressed in HSCs and meg precursors, prime candidate chromosome 21 gene for TAM phenotype – upregulated in DS-AMKL cells and megakaryoblastic cell lines (Rainis, <i>et al</i> 2005)
Bach-1	Chromosome 21 gene Meg differentiation Erythroid differentiation	Upregulated in 2 independent DS-AMKL microarrays <sup>1,2</sup> . Rapidly upregulated on GATA-1 rescue of G1E cells <sup>3</sup>	Candidate chromosome 21 gene for TAM phenotype. High levels block terminal megakaryocytic differentiation, negative regulator of erythroid differentiation (Toki, <i>et al</i> 2005).
NFE2	Meg differentiation Erythroid differentiation	Rapidly upregulated on GATA-1 rescue of G1E cells <sup>3</sup>	Identified as erythroid transcription factor but knock-out mice have mild erythroid phenotype with complete absence of platelets (Shivdasani 2001). Likely to be direct transcriptional target of GATA-1 and may be differentially regulated with GATA-1s
Ski	Meg differentiation	None	Selective expression in Meg-erythroid progenitors with down regulation required prior to terminal differentiation. Prevents GATA-1 DNA binding via direct interaction with C-zinc finger (Ueki, <i>et al</i> 2004). Should act as a marker for MEPs and may be differentially regulated by GATA-1s
Gp1b (platelet glycoprotein 1b)	Meg differentiation	Not properly upregulated in transgenic murine GATA-1s megakaryocytes <sup>4</sup>	Important for meg proliferation and maturation – both defective in GATA-1s mutants. Down-regulation associated with cellular transformation (Li, <i>et al</i> 2008b). Identified by microarray as possible GATA-1s target
D6	GATA-1 target gene, Meg differentiation	None	Transcription known to be regulated by GATA-1 (McKimmie, <i>et al</i> 2008), high expression in megakaryocytes.
GATA-1	Meg differentiation Erythroid differentiation	Upregulated in DS-AMKL vs non-DS-AMKL <sup>1</sup>	GATA-1 regulates its own expression. Assay measures human GATA-1 (transgene murine) therefore allows assessment of effect of transgene on endogenous GATA-1 levels

EKLF (erythroid krüppel like factor)	Erythroid differentiation	Rapidly upregulated on GATA-1 rescue of G1E cells <sup>3</sup>	Known to restrict megakaryocytic differentiation and promote erythroid differentiation (Bouilloux, <i>et al</i> 2008). GATA-1 target gene.
Zeta-globin	Erythroid differentiation		Marker of embryonic erythropoiesis, may reveal ontogeny related differences in GATA-1 isoform effects
B-globin major	Erythroid differentiation	Rapidly upregulated on GATA-1 rescue of G1E cells <sup>3</sup>	Marker of adult erythropoiesis to compare with zeta-globin above
PU.1 (spi-1)	Lineage Choice	Not properly repressed in transgenic murine GATA-1s megakaryocytes <sup>4</sup>	PU-1 and GATA-1 directly interact and determine lineage choice (myeloid/granulocytic versus meg- erythroid) (Rekhtman, <i>et al</i> 1999).
Ikaros	Lineage Choice	Not properly repressed in transgenic murine GATA-1s megakaryocytes <sup>4</sup>	As above, known to interact with GATA-1 and determine lineage choice (lymphoid versus myeloid) but also important later for $\gamma$ to $\beta$ -globin switching (Bottardi, <i>et al</i> 2008).
GATA-2	Lineage Choice	Not properly repressed in transgenic murine GATA-1s megakaryocytes <sup>4</sup>	Important for HSC maintenance but also higher levels needed for megakaryocytic differentiation and lower levels for erythroid differentiation (Ikonomi, <i>et al</i> 2000).
Cyclin D2 (CCND2)	Cell cycle Cell survival	Not properly up- regulated in transgenic murine GATA-1s megakaryocytes <sup>4</sup>	Cyclin family regulated by GATA-1 and important for megakaryocyte growth and terminal differentiation via regulation of cell cycling (Matsumura, <i>et al</i> 2000).
c-myb	Cell cycle Cell proliferation	Not properly repressed in transgenic murine GATA-1s megakaryocytes <sup>4</sup>	Essential for definitive haematopoiesis. High levels inhibit terminal differentiation. Target genes are involved in cell cycle progression and survival pathways. (Sakamoto, <i>et al</i> 2006)

**Table 5.1 Genes included in qPCR expression assays**

Abbreviations: Meg=megakaryocytic <sup>1</sup>(Bourquin, *et al* 2006), <sup>2</sup>(Ge, *et al* 2006), <sup>3</sup>(Welch, *et al* 2004) <sup>4</sup>(Li, *et al* 2005). Colours refer to scheme used to display microarray gene results i.e. blue = megakaryocytic differentiation, pink = erythroid differentiation, green = lineage choice, yellow = cell cycle

all RNA extractions for a sample group (e.g. K562 GATA-1FL/GATA-1s/empty vector) were performed on the same day. Reverse transcription into cDNA was again performed in one batch per sample group using Affinity Script reverse transcriptase as described in section 2.8.4. 2µg starting RNA were used and cDNA samples were then diluted 1 in 10 before use in the TLDA plate assay.

A number of different analyses were performed using different calibrators i.e.

1. An assessment of baseline gene expression differences between K562 and Meg-01 cells (using empty vector transfected cells) with K562 empty vector as the calibrator.
2. A comparison between empty vector, full-length and short transgenes in K562 cells, using K562 empty vector as the calibrator.
3. A comparison between empty vector, full-length and short transgenes in Meg-01 cells, using Meg-01 empty vector as the calibrator.

#### **5.4.7.1 K562 vs. Meg-01 gene expression**

Initially baseline gene expression differences were examined between the two cell lines K562 and Meg-01 by analysing the results of gene expression from 3 Meg-01 empty vector control samples versus 3 K562 empty vector control samples. It should be noted that this kind of comparison between different cell types (rather than within a single cell line) is only valid if these cells express similar amounts of the endogenous control gene per unit of cDNA. If not, then apparent differences in gene expression will reflect changes in the baseline of the control gene rather than changes in the test gene of interest. To test this, the CT at which the endogenous control (18SrRNA) crossed the threshold was compared for standardised amounts of K562 (3 transgenes with 3 samples per transgene) and Meg-01 (3 transgenes with 3 samples per transgene) template (i.e. 2µg starting RNA, reverse transcribed in one batch and resultant cDNA diluted 1 in 10 before use), results are shown in Fig 5.14a. Differences in levels of expression of 18SrRNA between the two cell lines would be manifested as a consistent trend in 18SrRNA expression, either higher or lower with one cell line. Instead a small degree of random variability which was not statistically significant was seen. Therefore these two cell lines can be directly compared. The level of gene expression from the K562 samples were arbitrarily set at 1 (i.e. the K562 samples were used as the calibrator for fold change) and compared to levels of expression (expressed as fold-change in gene expression) in the Meg-01 cells. Results are shown in

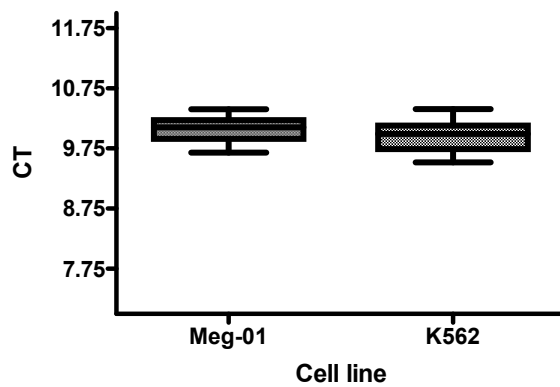


Fig 5.14b. Groups of genes are colour coded to emphasise their particular roles in megakaryocytic differentiation (blue), erythroid differentiation (pink), lineage choice (green) and cell cycle control (yellow) – many genes obviously fall into more than one category and have been assigned a group according to their most relevant role. Statistical analysis was performed by comparing two sample groups using an unpaired two-tailed Student's t-test (assuming unequal variance); values of  $p \leq 0.05$  were considered significant. Significant differences are shown on the graphs using standard nomenclature (\* =  $p \leq 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ ).

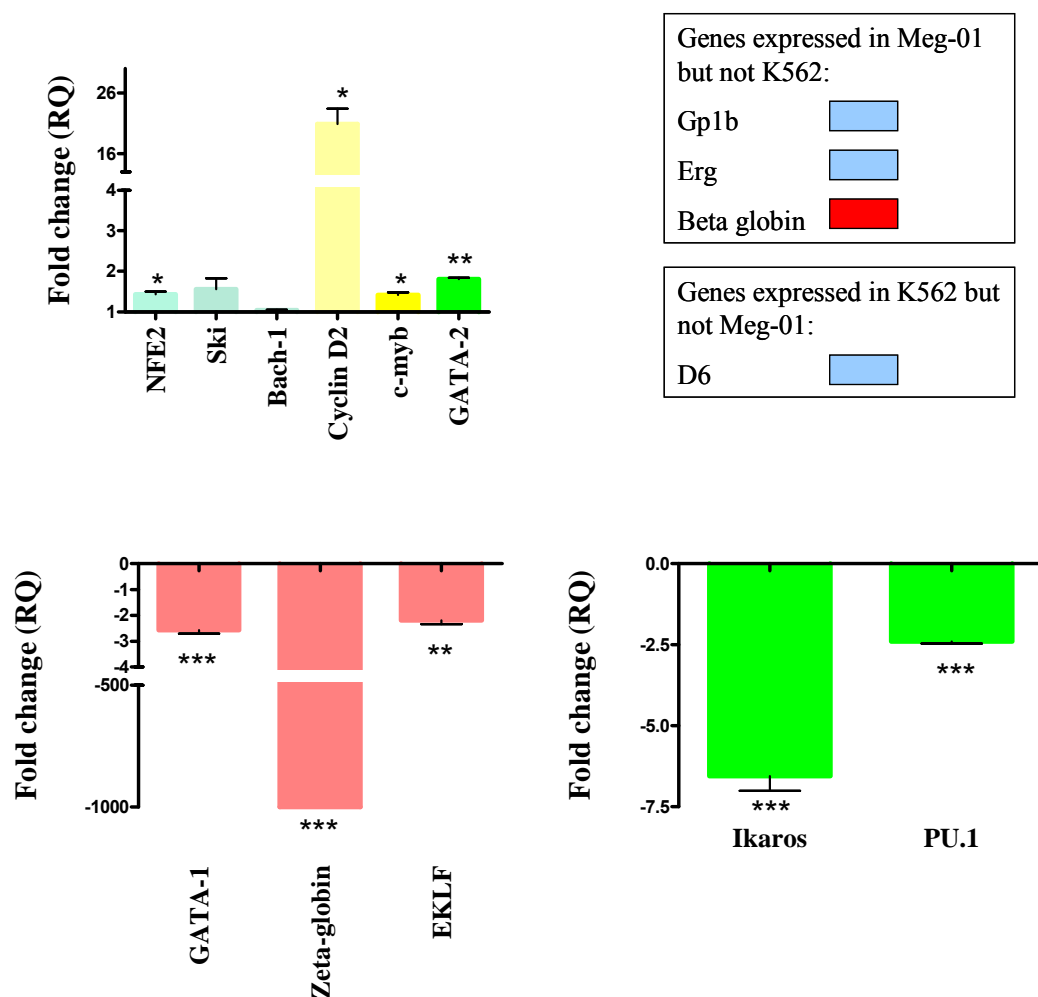
From Fig 5.14b it can be seen that Meg-01 cells have higher baseline levels of expression of the megakaryocytic transcription factor NFE2 ( $p=0.02$ ) and genes important for proliferation and survival of haematopoietic progenitors cyclin D2 ( $P=0.015$ ), c-myc ( $p=0.022$ ) and GATA-2 ( $P=0.002$ ). Genes associated with alternative lineages- Ikaros and PU.1- are expressed at lower levels in Meg-01 than in K562, presumably reflecting the more committed state of the former. Bach-1 is not upregulated ( $p=0.10$ ) despite the presence of Trisomy 21 in these cells. Erg3, another chromosome 21 gene, and 2 other genes – Gp1b and  $\beta$ -Globin major - are expressed in Meg-01, but expression is not detectable in K562 cells, this makes calculation of relative expression levels impossible since the lack of a comparator makes RQ quantitation invalid.

In contrast K562 cells have higher baseline expression of erythroid genes – zeta globin ( $p < 0.0001$ ), EKLF ( $p=0.003$ ) and GATA-1 ( $p=0.001$ ) and lineage determining genes- Ikaros ( $P=0.0001$ ) and PU.1 ( $p=0.0003$ ). They also show expression of a GATA-1 target gene D6 which (like Gp1b, Erg3 and  $\beta$ -globin major in K562 above), is undetectable in Meg-01 cells making estimation of relative expression levels impossible.

(a) **Baseline CT for Meg-01 and K562  
stable Transfectants**



(b)



**Fig 5.14 Comparison of gene expression between Meg-01 and K562**

(a) 18SrRNA CT values for standardised amounts of K562 and Meg-01 template are not significantly different and therefore the two cell lines can be directly compared (b) Fold change (RQ) in Meg-01 cells (empty vector controls) compared to K562 cells (n=3) showing higher megakaryocytic and lower erythroid gene expression. \*=p≤0.05, \*\*=p<0.01 and \*\*\*=p<0.001

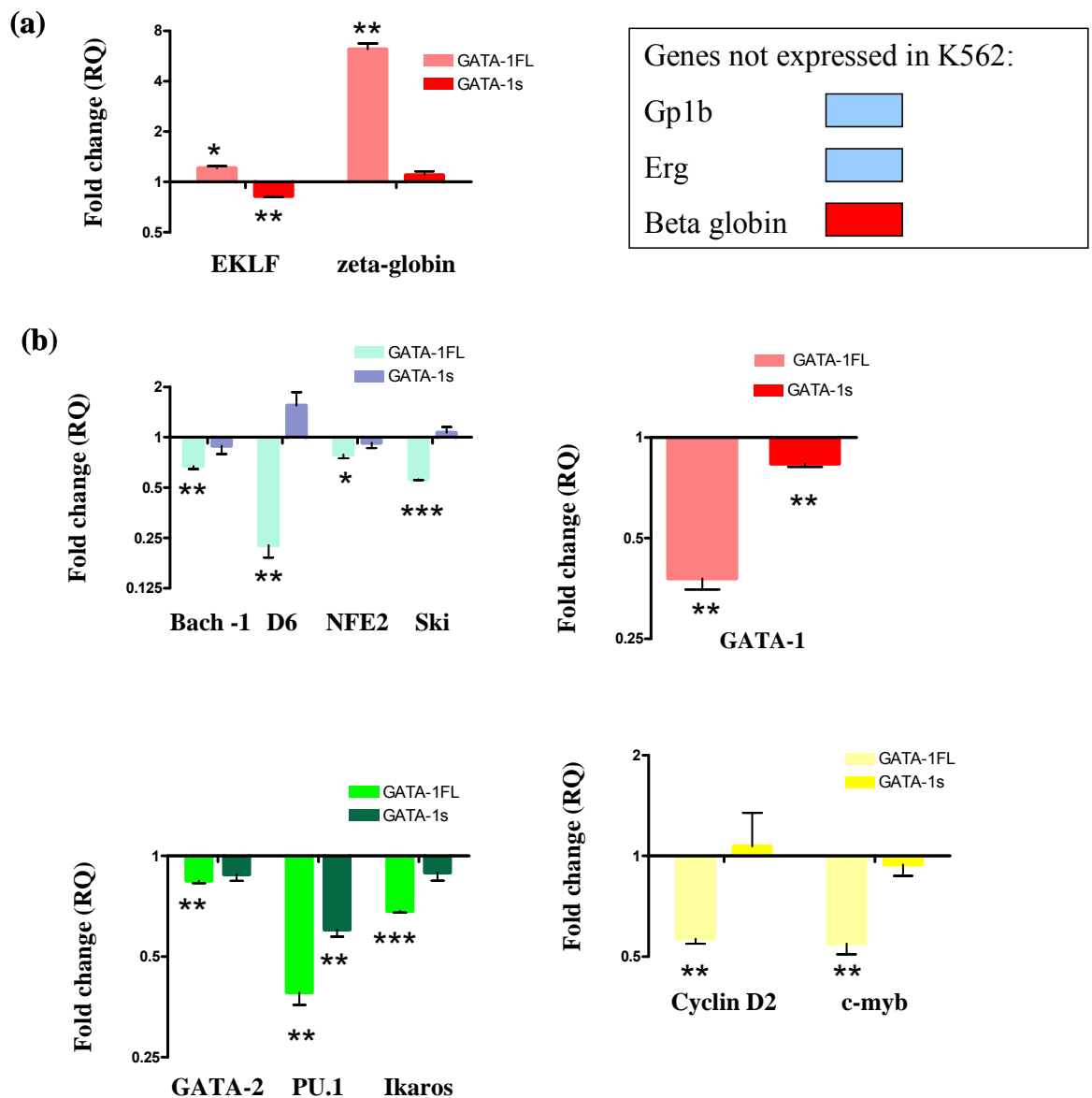
Overall these findings support the observations that the Meg-01 cell line has a number of features of megakaryocytic precursor cells, but still retains some features of more primitive haematopoietic cells (GATA-2, c-myb and cyclin D2). The lack of upregulation of Bach-1 is surprising, given the existing microarray data suggesting Bach-1 is consistently upregulated in AMKL (Bourquin, *et al* 2006, Ge, *et al* 2006), but may reflect the different stages of differentiation of these cells (adult haematopoietic precursors in CML vs. fetal haematopoietic precursors in TAM/AMKL). Unsurprisingly, K562 cells express lower levels of genes involved in megakaryocytic differentiation and higher levels of erythroid genes (presumably actively down-regulated in Meg-01 cells) when compared to Meg-01, their expression of PU.1 and Ikaros at higher levels than Meg-01 presumably reflects their more multipotent state, although recently a role for Ikaros in globin gene switching during erythroid maturation has been reported (Bottardi, *et al* 2008), so it may be acting as an erythroid gene in these circumstances. The expression of zeta-globin in K562 cells (an embryonic  $\alpha$ -globin chain) but not  $\beta$ -globin is intriguing. The production of fetal and embryonic haemoglobin in these cells has been reported previously (Baron and Maniatis 1986, Rutherford, *et al* 1979), despite their derivation from an adult patient with CML. It is tempting to speculate that the high levels of Ikaros may be relevant here in causing a reversion of these adult progenitors to fetal haemoglobin production. This hypothesis requires testing.

#### 5.4.7.2 The effects of GATA-1 transgene expression in K562 cells

Having established baseline characteristics of gene expression in the two cell lines, the effects of ectopic expression of the GATA-1 isoforms on this gene expression signature were examined. Data for K562 are presented in this section and Meg-01 in the section below. Three biological replicates from each cell line (K562 empty vector, K562 GATA-1FL and K562 GATA-1s) were cultured in 6-well plates for several passages and then harvested. Following RNA extraction and cDNA synthesis samples were run on TLDA plates as described above.

Fig 5.15 shows gene expression data displayed as fold-change in gene expression compared to levels in the K562 empty vector line (used as the calibrator with gene expression arbitrarily set at 1). Expression of Erg3,  $\beta$ -globin and Gp1b were again undetectable in all 3 transgenic cell lines precluding further analysis. All of the remaining 12 genes showed significant differences in gene expression levels between the K562 empty vector and the K562 GATA-1FL cell line (using a Student's two tailed t-test,  $p \leq 0.05$ ). The data are presented using a log 2 Y axis to allow fold reductions in gene expression to be

displayed with equivalent magnitude to fold increases in gene expression (i.e. from 1 to 0.5 = 2 fold reduction which is equivalent to a 2 fold increase):



**Fig 5.15 Alterations in gene expression profile with GATA-1 isoform over-expression in K562 cells**

**(a) genes showing significant up-regulation with the GATA-1FL isoform (and, in the case of EKL, down-regulation with GATA-1s) (b) Genes showing down-regulation with GATA-1 transgene expression** \*= $p \leq 0.05$ , \*\*= $p < 0.01$  and \*\*\*= $p < 0.001$

Genes showing significant up-regulation with GATA-1FL were the two erythroid specific genes Zetaglobin ( $p=0.01$ ) and, to a lesser extent, EKL ( $p=0.03$ ). Significant down-regulation was seen for genes associated with primitive haematopoietic precursors, cell cycling or other lineage choices: c-myb ( $p=0.008$ ), cyclin D2 ( $p=0.002$ ), GATA-2 (albeit a small fold change) ( $p=0.008$ ), Ikaros ( $p=0.0004$ ) and PU.1 ( $p=0.003$ ). In addition down-regulation of megakaryocytic genes was also seen: Ski ( $p=0.0004$ ), NFE2 ( $p=0.03$ ), D6

( $p=0.002$ ) and Bach-1 ( $p=0.006$ ). Finally, consistent with GATA-1 autoregulation GATA-1 levels were also significantly reduced ( $p=0.002$ ), note that the TLDA plates measure human GATA-1 and that the transgene is murine GATA-1 allowing effects of transgene expression on endogenous GATA-1 levels to be assessed. ***Overall, these findings are consistent with a dominant role of GATA-1FL in driving K562 cells along the erythroid lineage and away from a more multipotent state.***

In contrast, ectopic GATA-1s expression was unable to modulate gene expression to the same degree. Only 3 genes were significantly altered when compared to expression levels in K562 empty vector controls. All 3 genes were down regulated – PU.1 ( $p=0.004$ ), GATA-1 ( $p=0.01$ ) and EKLF ( $p=0.004$ ). Despite significant down regulation of PU.1 and GATA-1 this was to a lesser extent than seen with GATA-1FL expression (with a statistically significant difference between K562 GATA-1FL and K562 GATA-1s levels of expression of PU.1 ( $p=0.008$ ) and GATA-1 ( $p=0.0004$ ) when their fold-change values were directly compared). Other haematopoietic precursor genes (GATA-2, Ikaros) down-regulated in GATA-1FL cells were also down-regulated in GATA-1s, but to a lesser extent. Other genes (down-regulated by GATA-1FL) showed no evidence of repression by GATA-1s particularly the megakaryocytic genes – ski, NFE2, D6, Bach-1 and cell proliferation associated genes – c-myc and Cyclin D2. Finally the erythroid genes – zeta globin and EKLF, upregulated with GATA-1FL expression, showed failure of upregulation (Zeta globin) or active down-regulation (EKLF) with GATA-1s.

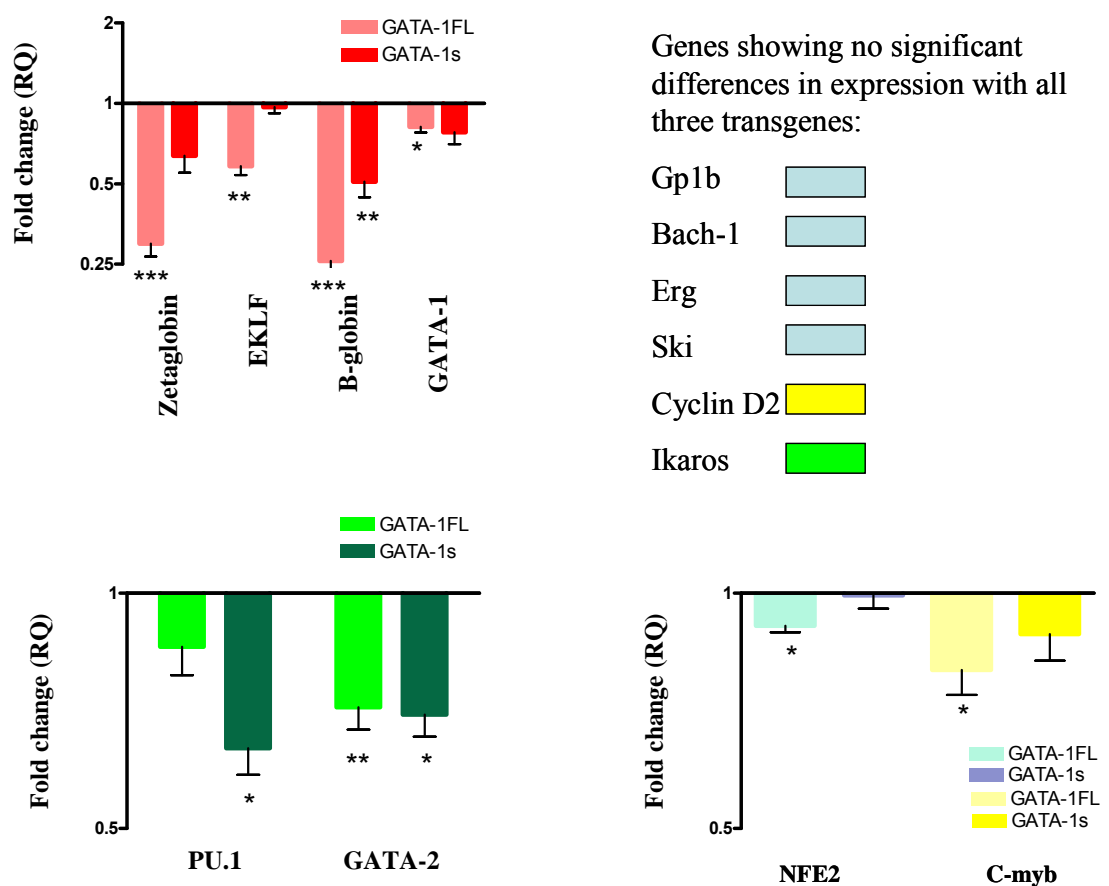
These findings suggest that GATA-1s expression in K562 cells is still able to promote differentiation away from haematopoietic precursor cells towards a meg-erythroid fate (by down-regulating genes associated with haematopoietic precursors) but that it does not drive erythroid commitment to the degree that GATA-1FL does. In addition the lack of repression of cyclin D2 and c-myc along with persistent expression of megakaryocytic genes are consistent with the proposed role of GATA-1s in the maintenance of a population of proliferative meg-erythroid precursors. ***Overall, this pattern of expression suggests a preference for megakaryocytic differentiation (lack of repression of megakaryocytic genes, failure to upregulate erythroid genes, active repression of EKLF) rather than erythroid differentiation with transgenic GATA-1s expression – the opposite of GATA-1FL.***

An alternative interpretation of the above is that these relative differences in GATA-1FL and GATA-1s effects on gene expression merely reflect differences in the expression levels of GATA-1 in the two cell lines i.e. the GATA-1FL construct is expressed at a

higher level than the GATA-1s construct. Although this cannot be discounted, the active down-regulation of EKLF expression with GATA-1s as opposed to up-regulation with GATA-1FL suggests that GATA-1s has a distinct functional role, at least with some target genes. In addition Meg-01 data discussed below - where the GATA-1s gene is expressed at higher levels than GATA-1FL - suggest these changes are not merely concentration dependent effects.

### 5.4.7.3 The effects of GATA-1 transgene expression in Meg-01 cells

Following sample preparation as described for K562 cells above, the effects of the two isoforms were also studied in Meg-01 cells. Results are shown in fig 5.16 and summarised below:



**Fig 5.16 Alterations in gene expression profile with GATA-1 isoform over-expression in Meg-01 cells**

\*= $p \leq 0.05$ , \*\*= $p < 0.01$  and \*\*\*= $p < 0.001$

In contrast to its effects in K562 cells, the GATA-1FL transgene expression produced a significant fall in the expression of erythroid associated genes – zeta-globin ( $p=0.0002$ ), EKLF ( $p=0.002$ ),  $\beta$ -globin ( $p<0.0001$ ) and GATA-1 ( $p=0.03$ ). Modest, but statistically significant, reductions in c-myb ( $p=0.03$ ), NFE2 ( $p=0.03$ ) and GATA-2 ( $p=0.008$ ) were also seen. The megakaryocytic genes Gp1b, Bach-1, Erg3 and Ski, along with cyclin D2, showed no significant changes in expression. Ikaros and PU.1 showed small reductions in gene expression but these were not statistically significant. It should be noted that the expression of these two haematopoietic precursor genes is already significantly down-regulated in Meg-01 cells compared to K562. In addition the megakaryocytic genes (with the exception of Bach-1) are already up-regulated in Meg-01 cells (as shown in Fig 5.14b). The degree to which additional GATA-1 isoform expression can alter these pre-existing patterns is unclear. Since GATA-1 plays important roles in both megakaryocytic and erythroid differentiation it is not surprising that GATA-1FL expression in a more committed progenitor might cause down-regulation of erythroid genes. As discussed in Chapter 1, GATA-1 forms complexes with multiple different partner proteins and it is the particular combination of proteins that determines repression versus activation of target genes (Rodriguez, *et al* 2005). It must be assumed that the protein partners of GATA-1FL differ in K562 and Meg-01 cells leading to a balance favouring erythroid gene activation in K562 and repression in Meg-01. ***In summary, it appears that the GATA-1FL isoform when ectopically expressed in megakaryocytic cells leads to down-regulation of erythroid genes but no further active upregulation of megakaryocytic genes.***

The GATA-1s isoform expressing cells produced a significant down-regulation of GATA-2 ( $p=0.02$ ),  $\beta$ -globin ( $p=0.01$ ) and PU.1 ( $p=0.02$ ). All other genes showed no significant change in gene expression. Again this is consistent with published results showing that GATA-1s expressing megakaryocytic precursors fail to adequately down-regulate c-myb, ikaros, cyclin D2, and GATA-1, although the significant down-regulation of PU.1 in GATA-1s and not GATA-1FL expressing cells is unexpected. Levels of gene expression may be relevant here: as seen in Fig 5.9 in K562 cells both GATA-1FL and GATA-1s show detectable expression on western blot with good levels of the murine GATA-1FL transgene (using murine specific N6 antibody), in contrast in Meg-01 cell lines increased expression of the GATA-1s protein is easily detectable by western blot whereas GATA-1FL expression is not seen using the N6 antibody (fig 5.9b). Therefore the effects of high level GATA-1s expression may have been able to down-regulate PU.1 to a greater extent than low level GATA-1FL expression producing this anomalous result. However, if this is the case then this does not explain the more profound down-regulation of erythroid genes with GATA-1FL compared to GATA-1s. It is interesting that, despite this higher level

GATA-1s expression, a much smaller subset of genes were down-regulated than with low level GATA-1FL expression. The preferential down regulation of PU.1 and GATA-2 support a role for GATA-1s in allowing initial lineage choice or maintaining the differentiated state without the ability to drive terminal differentiation. Again the down-regulation of erythroid ( $\beta$ -globin), but not megakaryocytic, genes by GATA-1s suggests a preference for megakaryocytic over erythroid differentiation.

Finally, these data can be compared with the cell surface marker data presented in section 5.4.6.3. The increase in GpA expression seen with the GATA-1FL transgene in K562 cells is mirrored by the significant increases in erythroid gene expression at the mRNA level. Similarly the lack of upregulation of GpA in the GATA-1s K562 transgenics is consistent with the lack of induction of erythroid genes. In Meg-01 cells there was no evidence of erythroid up-regulation at the mRNA level despite higher levels of cell-surface GpA in the GATA-1FL and GATA-1s transgenics at baseline (although this level of GpA expression was not correlated with GFP expression –high expressers actually showed reduced GpA expression). Given that this relationship between GpA, GATA-1 transgene expression, and expression of other erythroid genes at the mRNA level in Meg-01 cells, is complex, this assay may not be a good marker of cellular commitment in this cell line. It is possible that other factors drive aberrant GpA expression in Meg-01 cells. The CD41 data is equally complex. As discussed in section 5.5.1.1 below, despite its widespread use as a marker of megakaryocytic commitment, CD41 is also expressed earlier in haematopoietic differentiation. The alterations in levels of expression at the cell surface are not reflected in mRNA expression patterns of other megakaryocytic genes in either K562 or Meg-01 cells. This suggests that it cannot be used as a surrogate marker for megakaryocytic lineage commitment in these cell lines.

*In summary, gene expression profiles show that K562 and Meg-01 cells differ at baseline, with Meg-01 expressing megakaryocytic genes at higher levels than K562. K562 express higher levels of Ikaros and PU.1 and erythroid genes. These findings are consistent with the proposed cell of origin of these two cell lines, with K562 arising from a more multipotent precursor, whilst Meg-01 appears to possess features of an early bipotent meg-erythroid precursor (with a tendency towards a megakaryocytic fate). The additional presence of GATA-1 transgene expression leads to alterations in gene expression patterns. Whilst GATA-1FL is able to drive erythroid gene expression in K562 cells it leads to down regulation of these genes in Meg-01 cells, presumably reflecting the different levels of lineage commitment in these two cell lines. GATA-1s expression in both cell lines shows much less potent effects on the expression of most*



*genes (with the exception of PU.1 and GATA-2), even when expressed at higher levels than the GATA-1FL transgene (as seen in Meg-01 stable transfectants). This suggests that, in agreement with the data from sections 5.3 (murine ES cells) and 5.4.6 (cell maturation status), GATA-1s is unable to drive terminal erythroid differentiation. In addition, it appears to act, in some contexts, as a hypomorphic allele of GATA-1, with alterations in gene expression that mimic GATA-1FL but to a lesser extent.*

## **5.5 The use of K562 and Meg-01 cells to model megakaryocytic and erythroid differentiation**

The K562 cell line was established in 1970 from a pleural effusion of a patient with CML in blast crisis (Lozzio and Lozzio 1979). Although originally thought to arise from an undifferentiated granulocytic precursor cell (Lozzio and Lozzio 1979), later studies suggested that it actually had properties more akin to a meg-erythroid precursor (Andersson, *et al* 1979, Rutherford, *et al* 1979, Tabilio, *et al* 1983). Erythroid properties were suggested by the fact that resting K562 cells express the red cell antigen glycophorin, usually first detected in proerythroblasts (Andersson, *et al* 1979). Furthermore, treatment of K562 cells with the known erythroid inducing agent Haemin (which is a crystalline chloride of heme ( $C_{34}H_{32}N_4O_4FeCl$ )) results in the synthesis of embryonic globin chains and the cells become positive for haemoglobin (detected by benzidine staining) (Andersson, *et al* 1979, Rutherford, *et al* 1979). More recent reports suggest Haemin treatment induces gene expression changes suggestive of erythroid commitment including expression of the Band 3 gene - an essential component of the red cell membrane (Yamada, *et al* 1998). In addition to these erythroid properties of K562 it was also shown that these cells could be induced to differentiate along the megakaryocytic lineage by treatment with Phorbol diesters, more commonly known as tumour promoting agents (Siebert and Fukuda 1985). The mechanism of action of these compounds in promoting megakaryocytic differentiation is unknown although it is thought that at least part of their activity comes from activation of protein kinase C with consequent effects on cell signalling pathways (Nagata, *et al* 1996). The commonest, and most effective, phorbol diester is 12-O-tetradecanoylphorbol-13-acetate (TPA). Treatment of K562 cells with TPA results in a reduction in erythroid properties and acquisition of various megakaryoblastic features including expression of megakaryocytic cell surface markers, production of platelet peroxidase, an increase in cell volume and an increase in DNA ploidy (Alitalo 1990, Tetteroo, *et al* 1984). Expression of platelet derived growth factor (PDGF),

plasminogen activator (PA) and plasminogen activator inhibitor (PAI-1) have also been demonstrated following TPA treatment (Alitalo 1990, Ogura, *et al* 1988).

Meg-01 cells were also derived from a patient with CML blast crisis (Ogura, *et al* 1985). In contrast to K562 these cells show clear megakaryoblastic characteristics in their resting state. During culture a sub-population of the cells become loosely adherent to the tissue culture dish and extend pseudopodia, these cells have even been reported to shed platelets into the culture media (Mroske, *et al* 2000). Again, treatment of these cells with TPA leads to further megakaryocytic differentiation with up-regulation of platelet factor 4 (PF4) and cyclo-oxygenase (Cox1), an increase in DNA ploidy and increased expression of platelet surface antigens (Mroske, *et al* 2000, Ogura, *et al* 1988). There is little published evidence on Haemin treatment of Meg-01 cells although one report suggests that these cells express globin genes at the mRNA level and levels of mRNA can be increased by Haemin treatment (Morle, *et al* 1992). In addition this same paper showed that 1/3 of Meg-01 cells became benzidine positive (indicating the presence of haemoglobin) after 3 days of culture with haemin. It seems likely, therefore, that these cells possess some residual erythroid potential.

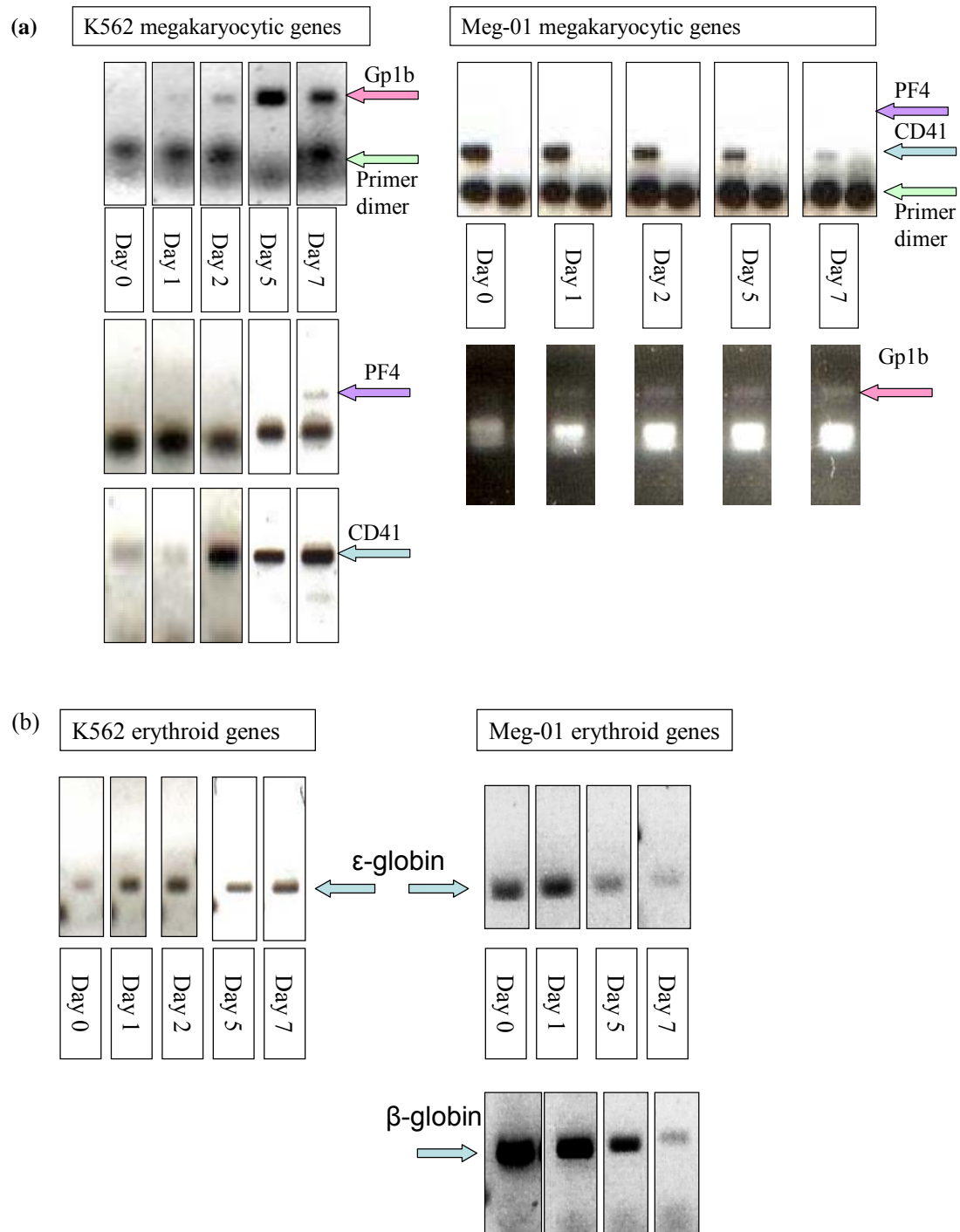
The following sections describe experiments using Haemin and TPA to examine the effect of GATA-1 isoform over-expression on the ability of K562 and Meg-01 cells to differentiate within the erythroid and megakaryocytic lineages. Differentiation status was assessed using ploidy determination, surface marker expression and gene expression profiling, as described above. Since TPA and Haemin differentiation assays were not in routine use in our laboratory, it was necessary to optimise and validate the experimental methods before assessing the impact of transgene expression. This is described below:

### **5.5.1 Validation of model systems**

#### **5.5.1.1 TPA**

Cells were treated with 100nM TPA as described in Materials and Methods section 2.13.3 and harvested after 6 hours or after 1, 2, 3, 4, 5 or 7 days, depending on the assay being used. Cells were noted to rapidly adhere to the culture dish a few hours after treatment and by 24 hours many cells with long pseudopodia were evident. This was particularly prominent with K562 cells; Meg-01 cells also adhered but remained rounder with few pseudopodia. These adherent cells, in both cell lines, then gradually detached over the following days. This phenomenon has been reported in the literature (Ogura, *et al* 1988).

To look for expression of megakaryocytic genes at the mRNA level, standard RT-PCR was performed on cDNA samples from days 0, 1, 2, 5 and 7 looking for evidence of induction of Platelet Factor 4 (PF4), CD41 (GpIIb) and Gp1b expression, as shown in Fig 5.17a.



**Fig 5.17 RT-PCR for megakaryocytic and erythroid genes following TPA treatment of Meg-01 and K562 cell lines**

(a) RT-PCR for megakaryocytic genes Gp1b, PF4, CD41 in K562 and Meg-01 cell lines on days 0,1,2,5, and 7 following TPA treatment (b) RT-PCR for erythroid genes  $\epsilon$ -globin and  $\beta$ -globin in K562 and Meg-01 cell lines on days 0,1,2,5 and 7 following TPA treatment

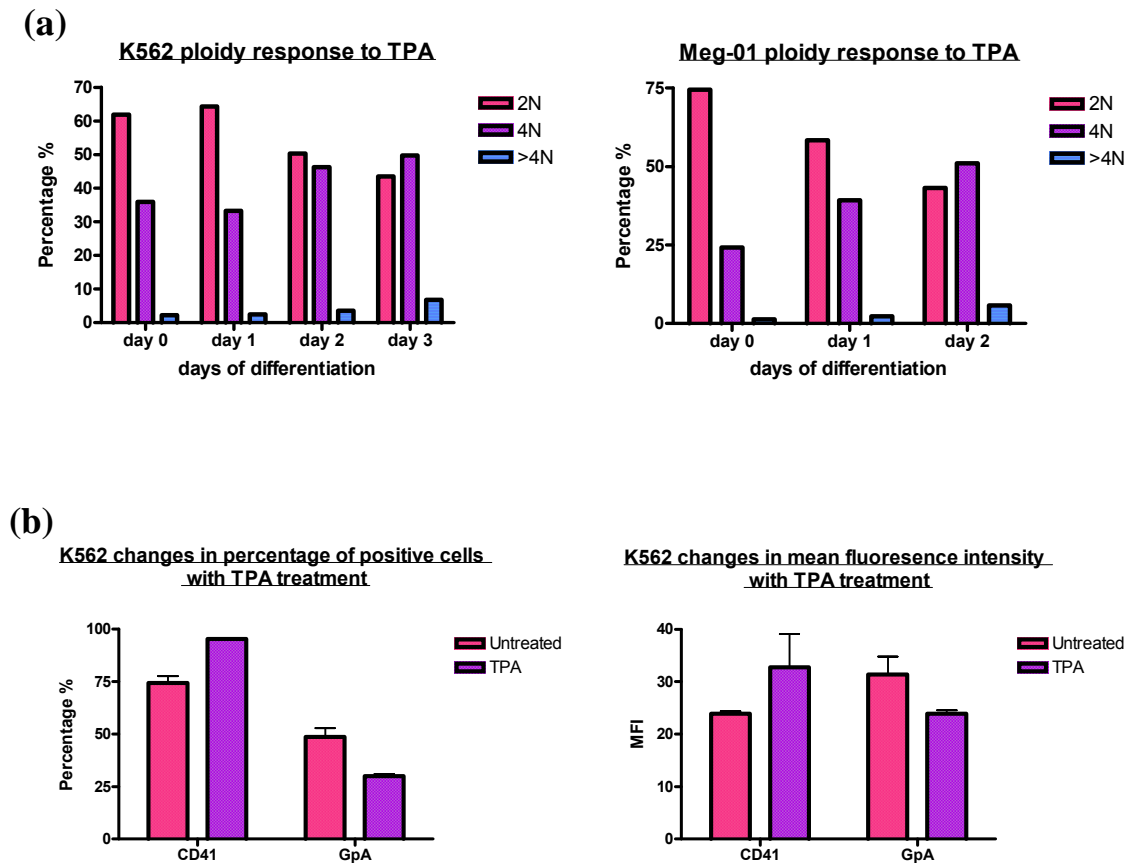
**1. The effect of TPA on megakaryocytic gene expression in K562 cells:** TPA treatment resulted in clear induction of Gp1b expression, which appeared maximal at day 5 (although it should be noted that this PCR is not optimised for quantitation). PF4 was induced at day 7 with undetectable bands before this. CD41 was weakly expressed at day 0 with an initial reduction in expression on day 1 followed by increasingly intense bands to day 7.

**2. The effect of TPA on megakaryocytic gene expression in Meg-01 cells:** Again induction of Gp1b was seen with undetectable expression at day 0 and low level expression from day 1 onwards increasing to day 7. Interestingly CD41 expression was high initially but progressively fell with subsequent days of differentiation. This latter result may reflect a role for this molecule earlier in the commitment to the megakaryocytic lineage, as discussed below. PF4 expression could not be detected in Meg-01.

The expression of erythroid genes was also examined to see if these fell with differentiation, Meg-01 cells down regulated both adult ( $\beta$ -globin) and embryonic globins ( $\epsilon$ -globin) on TPA induction as expected. However, unexpectedly K562 cells showed increased levels of  $\epsilon$ -globin expression (Fig 5.17b),  $\beta$ -globin expression was undetectable, consistent with the published reports of lack of adult globin expression in this cell line (Rutherford, *et al* 1979), this finding is discussed below.

As an independent measure of megakaryocytic differentiation, the effect of TPA treatment on DNA ploidy was assessed. Results are shown in Fig 5.18a. As can be seen from the figure, treatment of either cell line with TPA induced an increase in DNA ploidy from a modal 2N to modal 4N, with an increase in 8N peaks over time. Meg-01 acquired modal 4N ploidy by day 2, K562 by day 3, after TPA treatment. These changes are consistent, in magnitude and time course, with published reports in the literature of the DNA ploidy response of both cell lines to TPA (Jung, *et al* 2007, Ogura, *et al* 1988).

Finally, the levels of surface CD41a (GpIIb/IIIa) expression were assessed by FACS. Fig 5.18b shows that the percentage of CD41a<sup>+</sup> cells increased from around 75% to 95% and that the mean fluorescence intensity of these positive cells increased in K562 cells. In Meg-01 cells a similar rise in CD41a<sup>+</sup> cells was seen (data not shown). For comparison, expression of the erythrocytic cell surface antigen GpA was also assessed. In K562 expression fell with TPA treatment (fig 5.18b) whilst in Meg-01 it remained stable.



**Fig 5.18 The effect of TPA treatment on DNA ploidy and cell surface marker expression**

**(a) Treatment with TPA is associated with increases in DNA ploidy in both K562 and Meg-01 cell lines (n=1) (b) TPA treatment is also associated with increased intensity of CD41 expression and down-regulation of GpA expression by FACS analysis (n=3).**

CD41 was initially thought of as a specific megakaryocytic marker as it comprises the GpIIb component of the GpIIb/IIIa complex, which is essential in platelet adherence to sites of injury via its binding to fibrinogen. However, more recently two groups have reported that CD41 expression can be detected early during murine ontogeny at the time of emergence of definitive erythropoiesis in the yolk-sac and in vitro in embryoid bodies (day 6) (Mikkola, *et al* 2003, Mitjavila-Garcia, *et al* 2002). It then appears to be down-regulated in haematopoietic progenitors by the fetal liver stage. Further in vitro differentiation of embryoid body derived ES cells showed that CD41 could be switched on again later in development, presumably reflecting commitment to more mature lineages. Further support for a role in establishment of definitive haematopoiesis is supported by gene targeting experiments where expression of Cre-recombinase was linked to regulatory elements of the CD41 gene. This resulted in expression of a Cre-inducible reporter gene restricted to haematopoietic tissues but found in all myeloid and lymphoid cells suggesting CD41 expression is switched on in the earliest HSCs (Emambokus and Frampton 2003). CD41 expression has also been linked to the earliest stages of primitive erythropoiesis in

the yolk-sac (Ferkowicz, *et al* 2003). These observations are intriguing, particularly in view of the fact that CD41 knockout mice do not show any overt abnormalities in haematopoiesis apart from a platelet function defect (Tronik-Le Roux, *et al* 2000). This is also supported by human data – the disease Glanzmann's thrombasthenia is a severe autosomal recessive platelet function disorder characterised by mutations in either the GpIIb or GpIIIa genes. Affected individuals present with a severe bleeding disorder due to dysfunctional GpIIb/IIIa interaction with fibrinogen but do not show abnormalities in any of the other haematopoietic lineages (Saxena and Kannan 2008). Therefore, despite widespread expression, the role of CD41 in early haematopoietic precursors remains to be fully elucidated. From this discussion it can be seen that CD41 expression could be due to either terminal megakaryocytic differentiation or a residual marker of haematopoietic commitment. It is plausible that the differences in expression in K562 and Meg-01 reflect these dual roles for CD41.

When comparing K562 and Meg-01 cell lines it appears that Meg-01 more rapidly acquire megakaryocytic characteristics (i.e. more rapid increase in DNA ploidy, earlier expression of Gp1b, down regulation of globin gene expression) but the lack of PF4 expression and failure to repress GpA expression suggests that they retain some primitive meg-erythroid characteristics. In contrast K562 do express PF4 and repress GpA but show less repression of globin genes synthesis, again suggesting they retain some meg-erythroid characteristics. These findings mirror those reported in the literature which confirm the ability of both lines to differentiate into megakaryocytes but are often conflicting in their reporting of co-existent down regulation of erythroid markers. This is not surprising on two counts. Firstly there appears to be a marked tendency of both these cell lines, in the absence of overt selective pressure, to form subclones with distinct characteristics, particularly in the degree to which they display megakaryocytic or erythroid properties (Vainchenker, *et al* 1981). In fact, this observation was used as preliminary evidence for the existence of a bipotential meg-erythroid precursor, before later experiments (discussed in Chapter 1 section 1.1.2.2) confirmed the existence of these cells. Some of these studies may have used cell lines that had inadvertently (or deliberately in some cases (Ogura, *et al* 1988)) become skewed towards favouring megakaryocytic or erythroid differentiation, making comparison between results from different laboratories difficult. Secondly, these cells are derived from CML blast crisis and therefore are highly unlikely to display completely faithful lineage commitment, high levels of co-expression of supposedly lineage restricted markers have been seen, even after induction of differentiation *in vitro* (Tabilio, *et al* 1983). For the purposes of testing the hypothesis that GATA-1 isoforms may produce different effects on lineage commitment this lack of completely faithful differentiation is not a major obstacle,

but this caveat should be borne in mind especially when these systems are used as surrogate models for processes during physiological megakaryopoiesis.

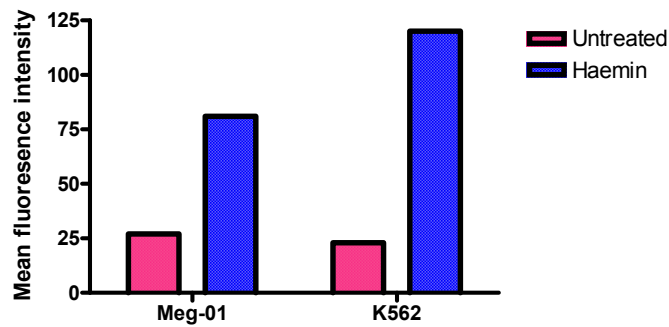
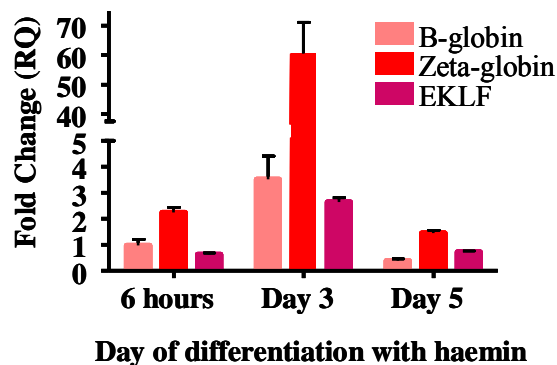
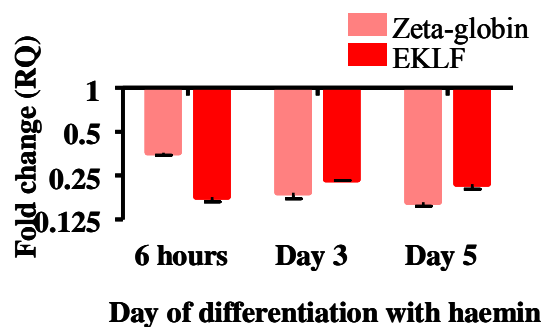
*In summary, as might be expected, Meg-01 and K562 do not behave identically when treated with TPA which probably reflects different stages of initial lineage commitment and cytogenetic differences between the two cell lines. However, these results confirm the ability of TPA to induce megakaryocytic differentiation in both Meg-01 and K562 cell lines as measured by expression of megakaryocytic genes, development of increased DNA ploidy and increases in the percentage of CD41+ cells and levels of surface CD41 expression.*

### 5.5.1.2 Haemin

Cells were cultured and treated with 10 $\mu$ M Haemin as described in chapter 2 section 2.13.4. Cells were harvested after 6 hours or 1, 2, 3, 4, 5, or 7 days differentiation according to the assay being performed. K562 cells were noted to form clusters in the first day after Haemin treatment and this persisted over the course of the differentiation. By day 2 of differentiation harvested K562 cells formed a pink cell pellet, reflecting haemoglobin synthesis and the colour of the pellet was noted to become progressively deeper pink/red over the course of the differentiation. This phenomenon was not seen in Meg-01 cells, initial clumping of cells was seen but cell pellets were not visibly pink.

To assess erythroid differentiation, in response to Haemin, levels of glycophorin A expression were assessed and quantitative PCR was performed for erythroid genes.

Meg-01 and K562 cells clearly up-regulated surface glycophorin A levels on haemin treatment as shown in Fig 5.19a. This was associated with significant upregulation of  $\beta$ -globin (4-fold) and the embryonic  $\alpha$ -globin chain - zeta-globin (64 fold) by day 3 of differentiation (fig 5.19b) in Meg-01 cells, as measured by qPCR. Interestingly K562 cells showed little evidence of an erythroid response in terms of gene expression (with down-regulation of zeta-globin and EKLF expression). This is despite clear haemoglobinisation of the harvested cell pellets from day 2 of differentiation onwards. However zeta-globin and EKLF expression were already higher in K562 cells (32 fold and 2 fold respectively) than Meg-01 so it may well be that the K562 response to Haemin is exerted at a post-transcriptional level or affects a different subset of genes responsible for terminal differentiation.

**(a) Glycophorin A expression with haemin treatment****(b) Meg-01 erythroid gene response to haemin****K562 erythroid gene response to haemin****Fig 5.19 Effects of haemin treatment**

**(a) Expression of glycophorin A (n=1) and (b) effect on gene expression patterns in K562 and Meg-01 cells (n=3) in response to haemin treatment**

These qPCR data illustrate some of the difficulties in interpreting fold-changes in gene expression rather than absolute quantitation of transcript levels, plus the limitations of studying gene expression purely at the mRNA level. However, for the purposes of examining whether the different GATA-1 isoforms show differential effects on induction of gene expression, fold-change is an adequate measure.



*Overall, the induction of haemoglobin production in K562 (as measured by development of pink/red colouration of the cell pellet), the induction of GpA cell surface expression in both cell lines, along with increased zeta-globin and EKLF expression in Meg-01 were taken to indicate that erythroid differentiation was induced by haemin in these two cell lines. Therefore, the impact of isoform expression on these parameters can be assessed. However the differences in qPCR data from the two cell lines suggest direct comparison between K562 and Meg-01 responses to haemin at the transcriptional level may be difficult.*

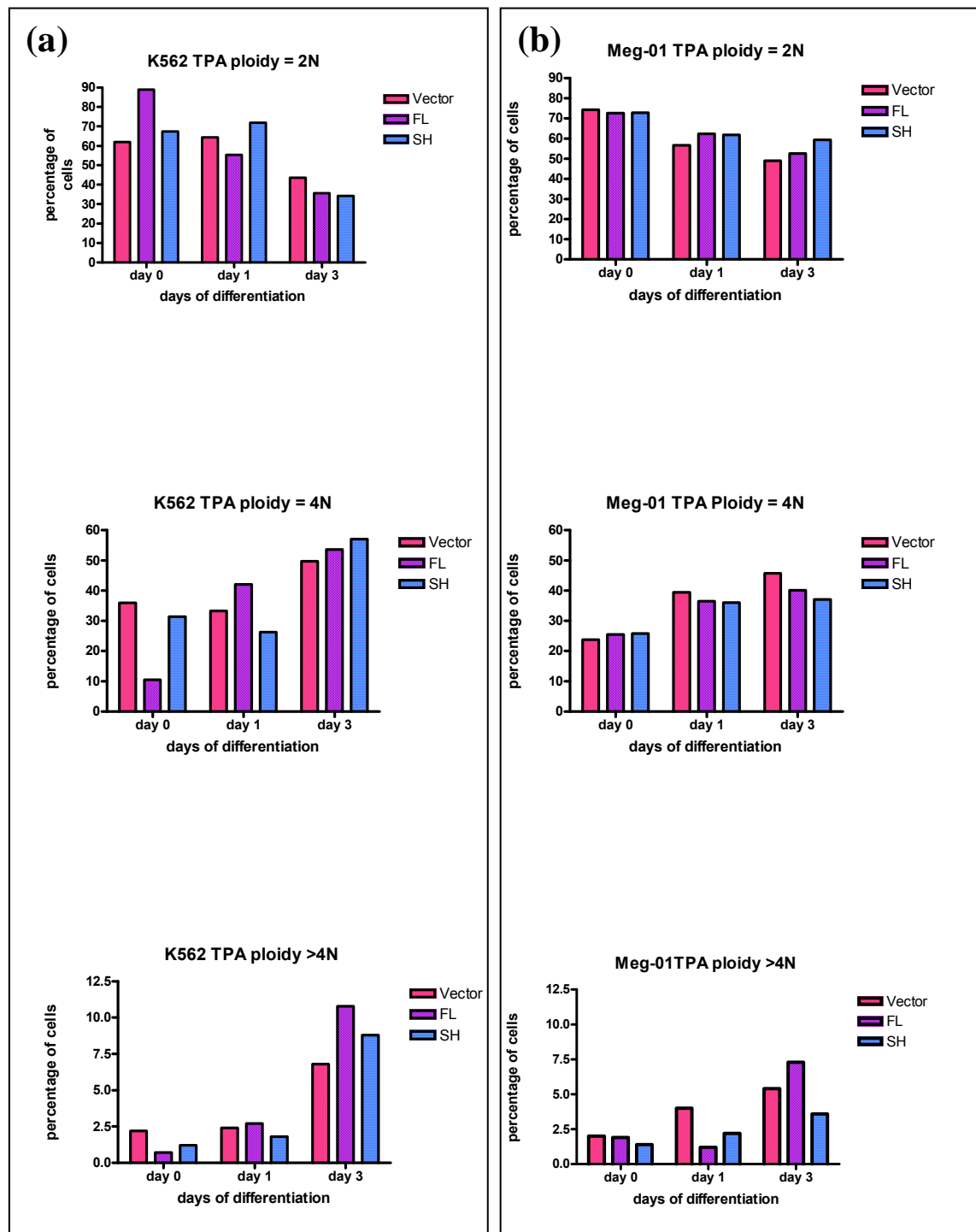
### **5.5.2 The effects of isoform expression on response to TPA in K562 and Meg-01 cells**

Having validated the ability of TPA to induce megakaryocytic differentiation in both K562 and Meg-01 cells the impact of GATA-1 isoform over-expression on this differentiation was assessed. Results are presented below:

#### **5.5.2.1 Effects on DNA ploidy**

Fig 5.20a shows effects on ploidy in K562 cells and Fig 5.20b in Meg-01 cells. Comparison between the two isoforms and the vector only control shows:

1. In K562 cells GATA-1FL expression is associated with lower baseline ploidy (90% of cells 2N compared to 60% in vector only control) but more rapid induction of 4N and >4N populations on TPA treatment suggesting accelerated differentiation down the megakaryocytic lineage. GATA-1s expression shows broadly similar baseline ploidy to the vector only control, but again a larger proportion of cells develop increased ploidy with TPA. In contrast to GATA-1FL the GATA-1s isoform produces more 4N cells and less >4N cells which may suggest it is capable of committing to megakaryocytic differentiation but that these cells mature less rapidly than with GATA-1FL.
2. In Meg-01 cells GATA-1FL expression does not show more rapid induction of 4N status but by day 3 of differentiation higher numbers of >4N cells are seen. GATA-1s expressing cells show lower numbers of both 4N and especially >4N cells compared to both GATA-1FL and empty vector by day 3.



**Fig 5.20 DNA ploidy changes in response to TPA**

(a) percentages of K562 transgenic cell lines with 2N, 4N and >4N DNA content as measured by propidium iodide incorporation on days 0, 1 and 3 following TPA treatment (representative data from 1 of 2 experiments shown), Vector =K562 empty vector, FL=GATA-1FL and SH=GATA-1s expressing cells (b) same analysis performed for Meg-01 cells (representative data from 1 of 2 experiments shown).

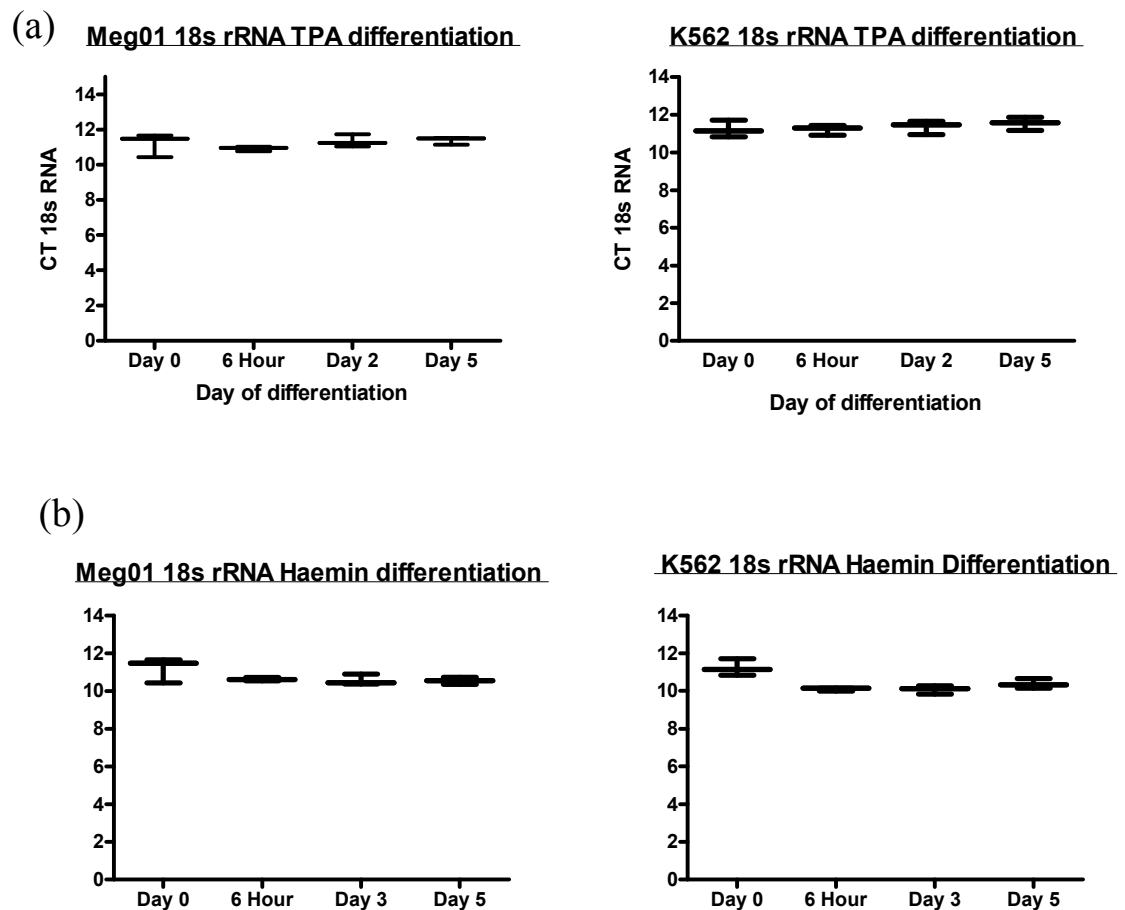
Two interesting observations emerge from these data. Firstly Meg-01 cells, despite clear baseline commitment to the megakaryocytic lineage, actually show lower degrees of ploidy on differentiation with TPA, regardless of the transgene expressed, when compared to K562 cells. Secondly, whilst GATA-1s is capable of driving megakaryocytic commitment

in K562 cells it appears to inhibit commitment (when compared to empty vector) in Meg-01. As discussed above, it is impossible to directly attribute these results to the presence of trisomy 21 in these cells. However, these data are certainly consistent with current theories suggesting that the combination of GATA-1s and trisomy 21 leads to outgrowth of immature megakaryocytic precursors blocked in differentiation, whilst cells carrying GATA-1s in the absence of trisomy 21 are capable of terminal megakaryocytic differentiation.

### 5.5.2.2 Effects on gene expression profiles

The effect of transgenic over-expression of the GATA-1 isoforms were compared to empty vector controls using TLDA plate qPCR analysis of cDNA extracted from day 0, 6 hour, day 2 and day 5 time-points. The 15 genes assayed were previously described in section 5.4.7. Prior to this analysis it was necessary to show that the expression of the reference housekeeping gene (18SrRNA) was not up or down-regulated with TPA treatment of cells as this would invalidate comparison between test genes at different time-points (levels of expression of the reference gene must remain constant if it is to accurately normalise for template cDNA levels between samples). Conventional Taqman qPCR was performed using a 1 in 10 dilution of cDNA prepared using 2µg starting RNA as described in materials and methods section 2.8.4. Each sample was assayed in triplicate and 3 biological replicates were run for each time point. A commercially available 18SrRNA primer and probe set (Applied Biosystems) was used. Results are shown in Fig 5.21a and statistical analysis (one way ANOVA) confirmed that levels of 18SrRNA expression are unaffected by TPA differentiation in both cell lines. Fig 5.21b presents the same analysis for haemin differentiation (discussed in section 5.5.3 below) and again confirms that levels of 18SrRNA expression are unaffected by Haemin treatment.

Having established that comparison between time-points was valid the effects of GATA-1 isoforms on TPA induced differentiation were analysed. Results are shown in Fig 5.22 (for K562) and Fig 5.23 (for Meg-01) and discussed below.

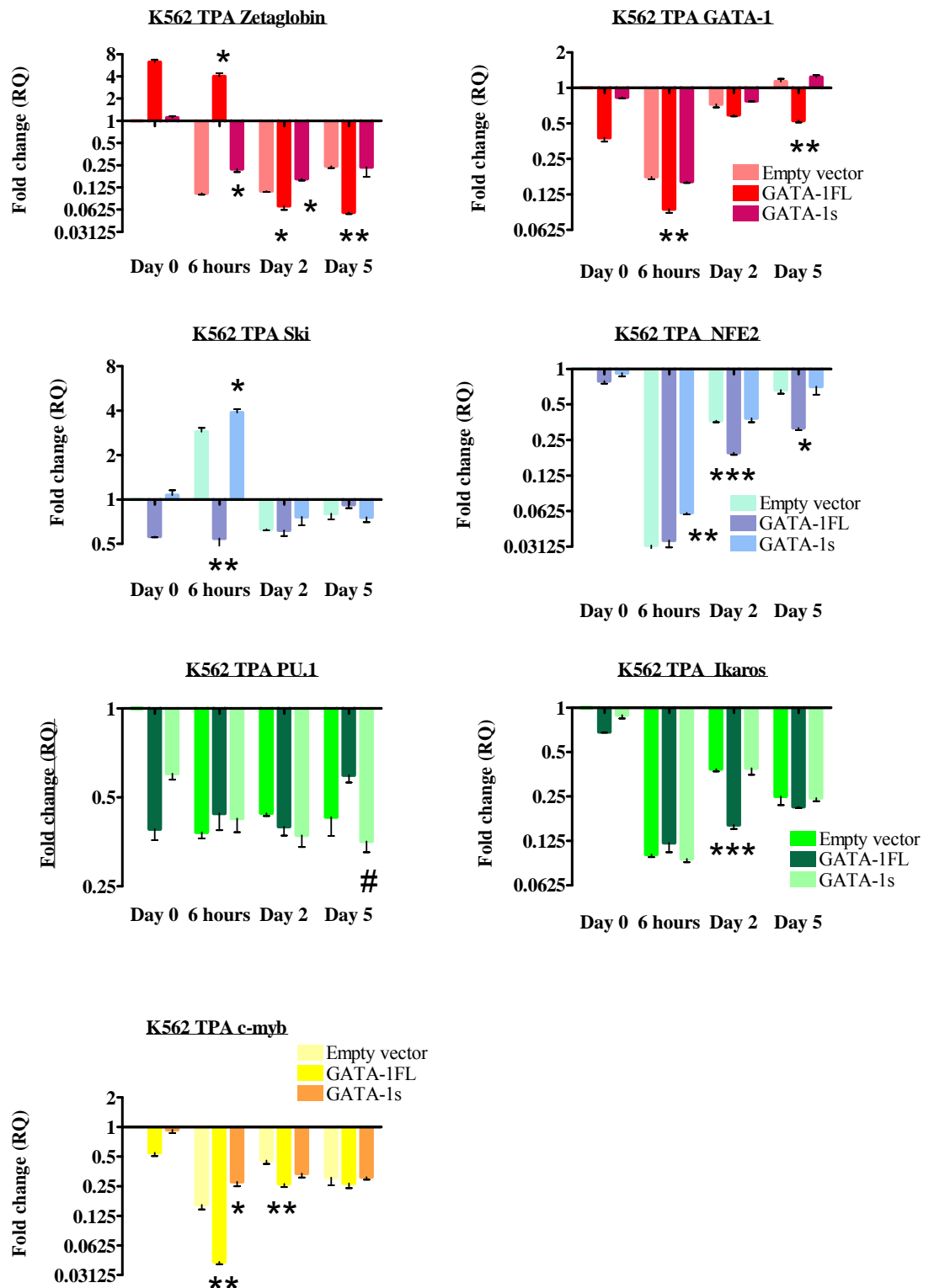


**Fig 5.21 Variation of 18SrRNA levels with TPA and Haemin treatment**

(a) with TPA treatment (b) with Haemin treatment. No significant differences were seen with CT values over time (one way ANOVA)

### 5.5.2.3 Gene expression in K562 transgenics

Fig 5.22 overleaf summarises the results of gene expression analysis in K562 transgenic cells when treated with TPA. All assayed genes (except 18SrRNA) showed statistically significant differences in gene expression over time when empty vector control cells were treated with TPA (i.e. day 0 versus 6 hours versus D2 versus D5 (one-way ANOVA looking at variation in gene expression levels in the K562 empty vector group)). To examine the impact of individual isoforms on these gene expression patterns, 2-way ANOVA analysis was performed with Bonferroni post-tests to compare replicate means for each time point. Untreated (day 0) empty vector control cells were used as the calibrator for fold change (i.e. their expression levels were arbitrarily set at 1.0).



**Fig 5.22 Gene expression patterns in TPA treated K562 transgenic cells**

Genes showing statistically significant variations in gene expression with the different GATA-1 isoforms are shown. Stars refer to the level of statistical significance when the test gene is compared between empty vector and the GATA-1FL or GATA-1s isoform \*= $p \leq 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . # = statistically significant difference when GATA-1FL and GATA-1s expression levels are directly compared ( $p < 0.05$ ). Genes are colour coded according to their dominant function. Red=erythroid differentiation, Blue=megakaryocytic differentiation, Green=lineage choice, Yellow=cell cycle/proliferation.

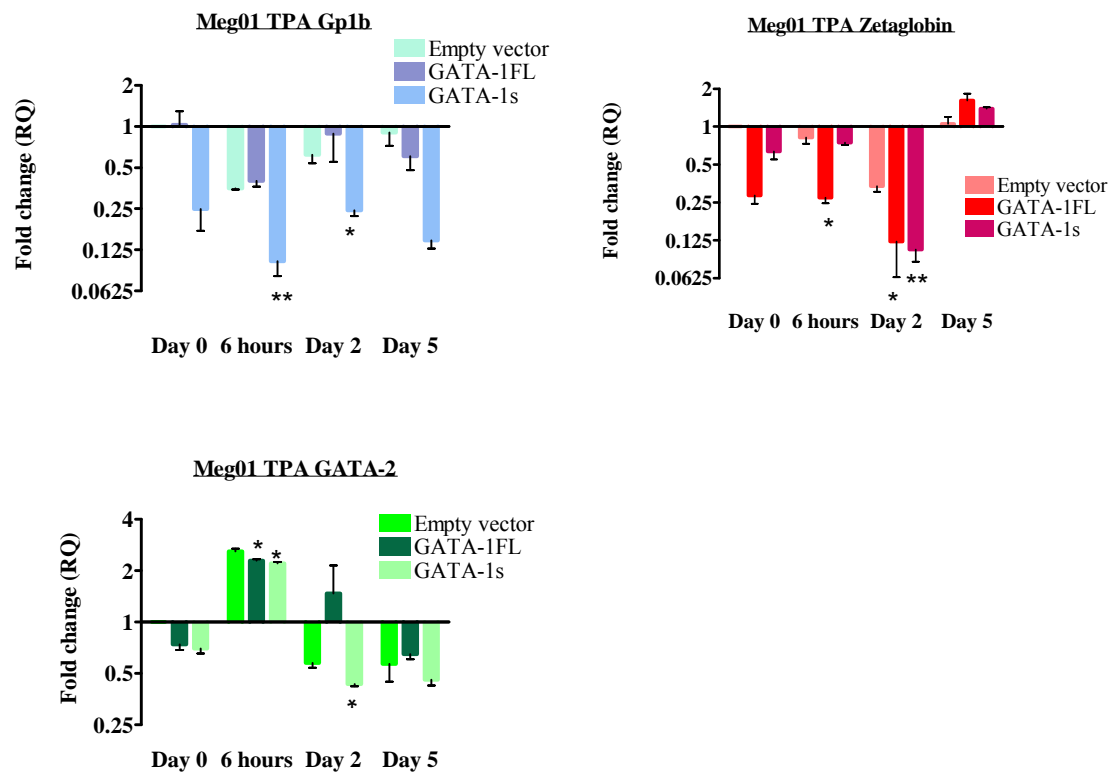
The GATA-1FL isoform produced statistically significant alterations in zetaglobin (delayed down-regulation at 6 hours, followed by more profound down-regulation at day 2 and 5), GATA-1 (down-regulation at 6 hours and day 5), ski (down-regulation at 6 hours), NFE2 (down-regulation at day 2 and 5) and Ikaro (down-regulation at day 2) expression levels when compared to the empty vector control. The GATA-1s isoform produced a significant change in ski gene expression (up-regulation at 6 hours) compared to empty vector and less repression of c-myc at 6 hours than empty vector or GATA-1FL. When comparing expression levels in GATA-1FL and GATA-1s transgenics, statistically significant differences in their effect on gene expression were seen for all the same genes and time points as seen with GATA-1FL vs. empty vector. One additional gene – PU.1 displayed a significant difference when expression levels were compared between the two isoforms, with GATA-1FL showing less down-regulation than with GATA-1s ( $p=0.05$ ). Overall this pattern of changes suggests that the GATA-1s isoform does not exert a dominant role on gene expression in the context of K562 TPA differentiation and instead largely mimics, with the notable exception of ski, the expression patterns seen in empty vector controls. In contrast GATA-1FL over-expression produces a number of dominant changes with higher levels of zetaglobin expression early on and lower levels of NFE2 later on. The higher levels of zetaglobin are explicable by the increased erythroid commitment of K562 cells expressing the GATA-1FL transgene (as discussed above), with the need to reverse this commitment prior to megakaryocytic differentiation. NFE2 has a more complex bimodal pattern of expression, is a GATA-1 target gene, and is important in both megakaryocytic and erythroid lineages. The lower levels of NFE2 at day 5 may simply reflect the lower levels of GATA-1 expression at this time-point. As far as GATA-1 expression is concerned the GATA-1FL expressing cells show greater levels of repression at all time points, this may reflect more effective autoregulation by the FL transgene than the GATA-1s transgene either by virtue of the presence of the N terminal domain or due to higher expression levels from the GATA-1FL construct.

Finally the expression patterns of the c-myc and ski genes are interesting. c-myc is a key haematopoietic transcription factor involved in stem cell self-renewal and lineage decisions and needs to be down-regulated to allow terminal differentiation. The reduced repression of c-myc with TPA treatment is similar to that observed for in vitro megakaryocytic differentiation of GATA-1s transgenic ES cells (Li, *et al* 2005) and may partly explain the apparent block to terminal maturation. Ski is an oncoprotein found to act in several signalling pathways resulting in transcriptional repression (Shivdasani 2001). Although widely expressed in haematopoietic and non-haematopoietic tissues at low level it shows a particular peak of expression in some murine bipotent meg-erythroid precursors with

much lower/undetectable levels in earlier precursors or unipotent megakaryocytic or erythroid precursors (Pearson-White, *et al* 1995). Other studies have established that ski expression is upregulated upon megakaryocytic differentiation of K562 cells with TPA (Namciu, *et al* 1994). In addition it has been shown to block erythroid differentiation by a direct interaction with GATA-1 involving the C-terminal Zinc finger preventing DNA binding of the GATA-1-ski complex to known GATA-1 target genes (Ueki, *et al* 2004). Overall these data suggest ski expression is essential for meg-erythroid commitment but that effective down-regulation is required to allow terminal differentiation. The higher levels of ski upregulation with the GATA-1s isoform may suggest an increase in bipotent MEP formation in these cells. The lack of upregulation of ski in GATA-1FL isoform expressing cells has several possible explanations. Since these cells clearly commit to the megakaryocytic lineage (as shown by ploidy analysis) this is unlikely to reflect decreased megakaryocytic differentiation in these cells, instead it is possible that the peak of expression occurred earlier (rapid modulation of mRNA levels has been documented 30 minutes to 3 hours following TPA treatment of K562 cells (Alitalo 1990)) and was over by 6 hours. Alternatively these cells may have already passed this ski peak prior to TPA treatment by virtue of their high baseline GATA-1FL expression. Further investigation of the role of ski in commitment to the meg-erythroid lineage is clearly needed and it would be extremely interesting to study ski levels in patients with TAM/AMKL.

#### **5.5.2.4 Gene expression in Meg-01 transgenics**

Next, the influence of GATA-1 isoform expression on TPA induced differentiation in Meg-01 cells was assessed. Similar to K562 cells, all studied genes with the exception of PU.1 showed a statistically significant difference in gene expression over time (i.e. day 0 versus 6 hours versus D2 versus D5 (one-way ANOVA looking at variation in gene expression levels in the Meg-01 empty vector group)). However, in contrast to K562 cells only a few genes showed differential regulation by the two isoforms. These results are shown in Fig 5.23. Expression of the GATA-1FL isoform produced a small but statistically significant change in GATA-2 levels (lower than empty vector at 6 hours of differentiation) and zeta-globin (down-regulated at the 6 hours and day 2). The GATA-1s isoform produced a similar modest down-regulation of GATA-2 at 6 hours and day 2, down-regulation of zeta-globin at day 2, and a more rapid and sustained down-regulation of Gp1b levels (significant at 6 hours and day 2). Otherwise all other genes showed similar responses in terms of alteration in gene expression with all three constructs.



**Fig 5.23 Gene expression patterns in TPA treated Meg-01 transgenic cells**

Genes showing significant alterations with GATA-1 isoform expression are shown. \*= $p \leq 0.05$ , \*\*= $p < 0.01$  and \*\*\*= $p < 0.001$

These results suggest that GATA-1 isoform over-expression has only a very modest or negligible effect on commitment to the megakaryocytic lineage in Meg-01 cells. The low level expression from the GATA-1FL transgene (as discussed above and shown in fig 5.9) makes interpretation of these results difficult although the modulation of zeta-globin expression levels suggests that even at this low level it is having some effect. The GATA-1s isoform appears to be expressed at higher level (fig 5.9) and its effects on Gp1b expression are intriguing – it may be that the excess GATA-1s in the transgenic cells is binding to Gp1b regulatory elements but cannot provide the necessary co-factors to stimulate expression leading to a down-regulation in these cells compared to the GATA-1FL and empty vector expressing cell lines i.e. it is acting in a dominant negative manner. It might be expected that Gp1b would show up-regulation during megakaryocytic differentiation but it may be that expression levels in the day 0 (i.e. comparator) cells are already high in this cell line and therefore no further up-regulation is required. An alternative explanation is that this assay is relatively inefficient and does not accurately reflect Gp1b levels – this is supported by a number of observations:



1. The high CT at which Gp1b expression crosses the threshold (which leads to difficulty in accurate relative quantitation since the number of cycles between detection of the endogenous control gene and the test gene is high- this magnifies any small errors in calculated efficiency values)
2. The high inter-plate variation in assay results (making many of the observed variations in Gp1b expression statistically insignificant).

For these reasons results obtained with Gp1b should be interpreted with caution.

*In summary, TPA treatment results in an increase in DNA ploidy which is more prominent in K562 cells than Meg-01. Expression of the GATA-1FL isoform causes increased numbers of cells with >4N DNA content in both cell lines. The GATA-1s isoform appears to inhibit the development of increased ploidy in Meg-01 cells (in the context of trisomy 21) whereas it mimics empty vector in K562 cells. Gene expression profiles show greater down-regulation of erythroid genes by the GATA-1FL isoform. The ski gene, a marker of bipotent MEPs, shows upregulation in K562 GATA-1s cells and down-regulation in K562 GATA-1FL cells. c-myb is inadequately repressed with GATA-1s and finally, Meg-01 GATA-1s transgenics show sustained and marked reduction in Gp1b mRNA levels. These latter observations suggest that expression of GATA-1s may block terminal differentiation and merit further investigation to determine their significance.*

### **5.5.3 The effects of isoform expression on response to haemin in K562 and Meg-01 cells**

The blast cells in TAM/AMKL show a mixed phenotype with both megakaryocytic and erythroid cell surface antigen expression and expression of erythroid and megakaryocytic genes at the mRNA level (Ito, *et al* 1995, Zipursky 2003). Therefore it is reasonable to hypothesise that the GATA-1s isoform may show unique properties in erythroid as well as megakaryocytic differentiation. To test this, the impact of GATA-1 isoform expression on in vitro erythroid differentiation was studied.

#### **5.5.3.1 Haemoglobin production in K562 cells**

As already described, Haemin treatment of K562 cells induces detectable haemoglobin as manifested by a pink colouration of the cell pellet on harvesting. This pink colouration

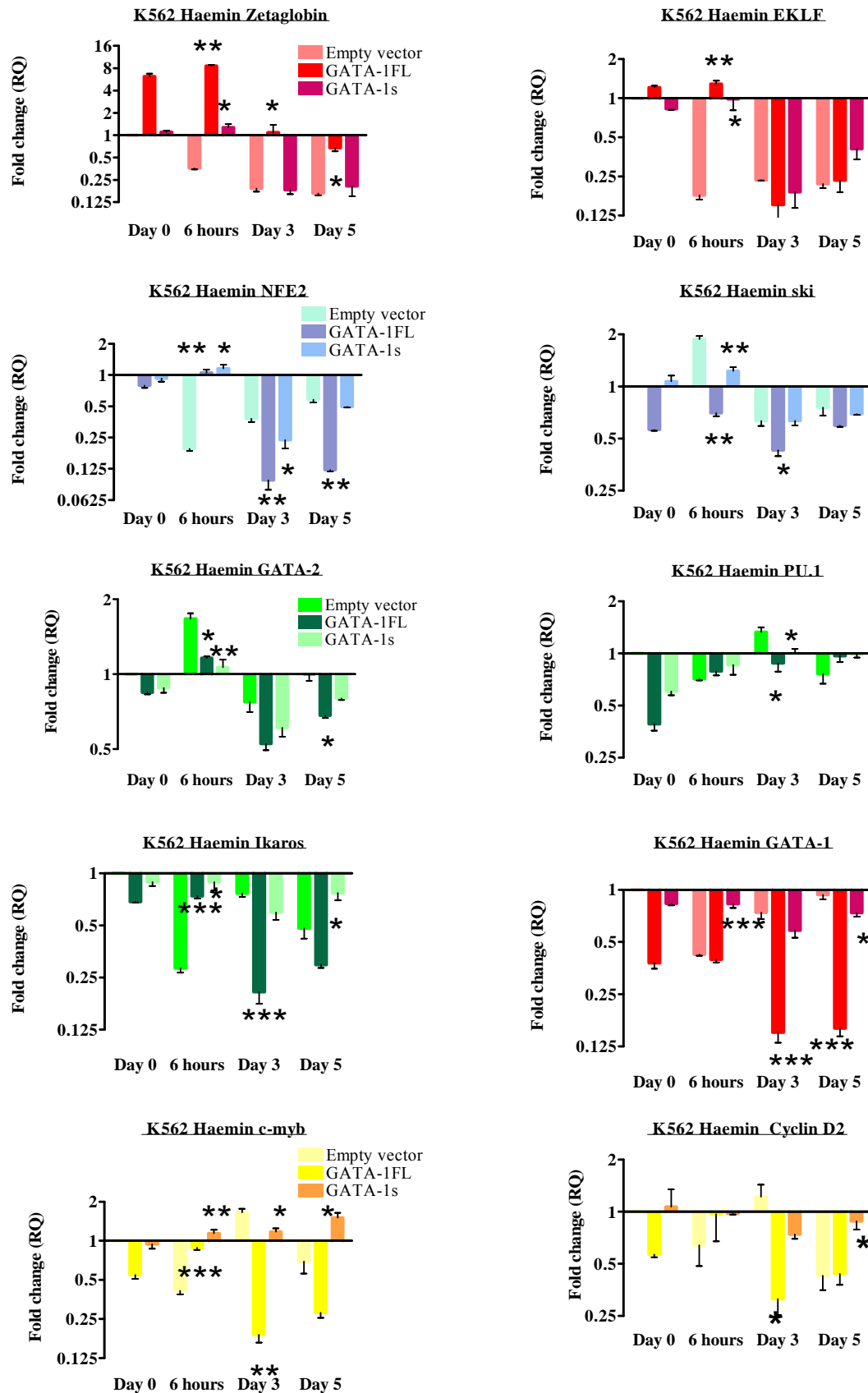
appeared earlier (day 1 compared to day 2), and was more intense, in K562 GATA-1FL transgenic cells. K562 empty vector and K562 GATA-1s cells both became pink on day 2 and showed similar intensities over the time course of TPA treatment (data not shown).

### 5.5.3.2 Gene expression in K562 transgenics

Following Haemin treatment cells were harvested at day 0, 6 hours, day 3 and day 5. RNA extraction and cDNA synthesis were performed as described above (section 5.4.7). TLDA plate assays were run on 3 biological replicates for each transgene at each time point. Results are shown in Fig 5.24 overleaf. Again all genes showed significant variations in expression with time during haemin differentiation. The majority of genes showed differential levels of expression with the 3 transgenes with only cyclin D2, Bach-1 and D6 showing no statistical differences between all 3 constructs.

1. Effect of the GATA-1FL isoform: GATA-1FL expression led to significant upregulation of zeta-globin expression (6 hours and day 3), EKLF expression (6 hours) and NFE2 expression (6 hours) compared to empty vector. Significant down-regulation of various known GATA-1 target genes was also seen – GATA-1 itself (day 3 and 5), GATA-2 (6 hours and day 5), Ikaros (6 hours and day 3), c-myb (day 3), NFE2 (day 3 and 5) and PU.1 (day 3). Ski again showed the most interesting pattern of expression with significant down-regulation in the GATA-1FL isoform at baseline and 6 hours and day 3 whilst the vector and short isoform initially upregulated ski at 6 hours and then down-regulated it by day 3. Since ski down-regulation is an essential event prior to terminal erythroid differentiation this rapid down-regulation with the GATA-1FL transgene may reflect the more rapid erythroid commitment seen in these cells.

2. Effect of the GATA-1s isoform: GATA-1s expression patterns generally mimicked GATA-1FL. Significant up-regulation (compared to empty vector control) was seen with zeta-globin (6 hours), EKLF (6 hours) and (unlike GATA-1FL) c-myb (6 hours, day 3 and day 5). Significant down-regulation was seen with GATA-1 (6 hours and day 5), GATA-2 (6 hours), Ikaros (6 hours and day 5), and ski (6 hours). However, when comparing GATA-1s levels to GATA-1FL levels of gene regulation, the FL isoform showed significantly larger effects (either more repression or more up-regulation) for zeta-globin, c-myb, GATA-1, Ikaros, NFE2 and Ski.



**Fig 5.24 Gene expression changes in K562 transgenics treated with Haemin**

\*= $p \leq 0.05$ , \*\*= $p < 0.01$  and \*\*\*= $p < 0.001$

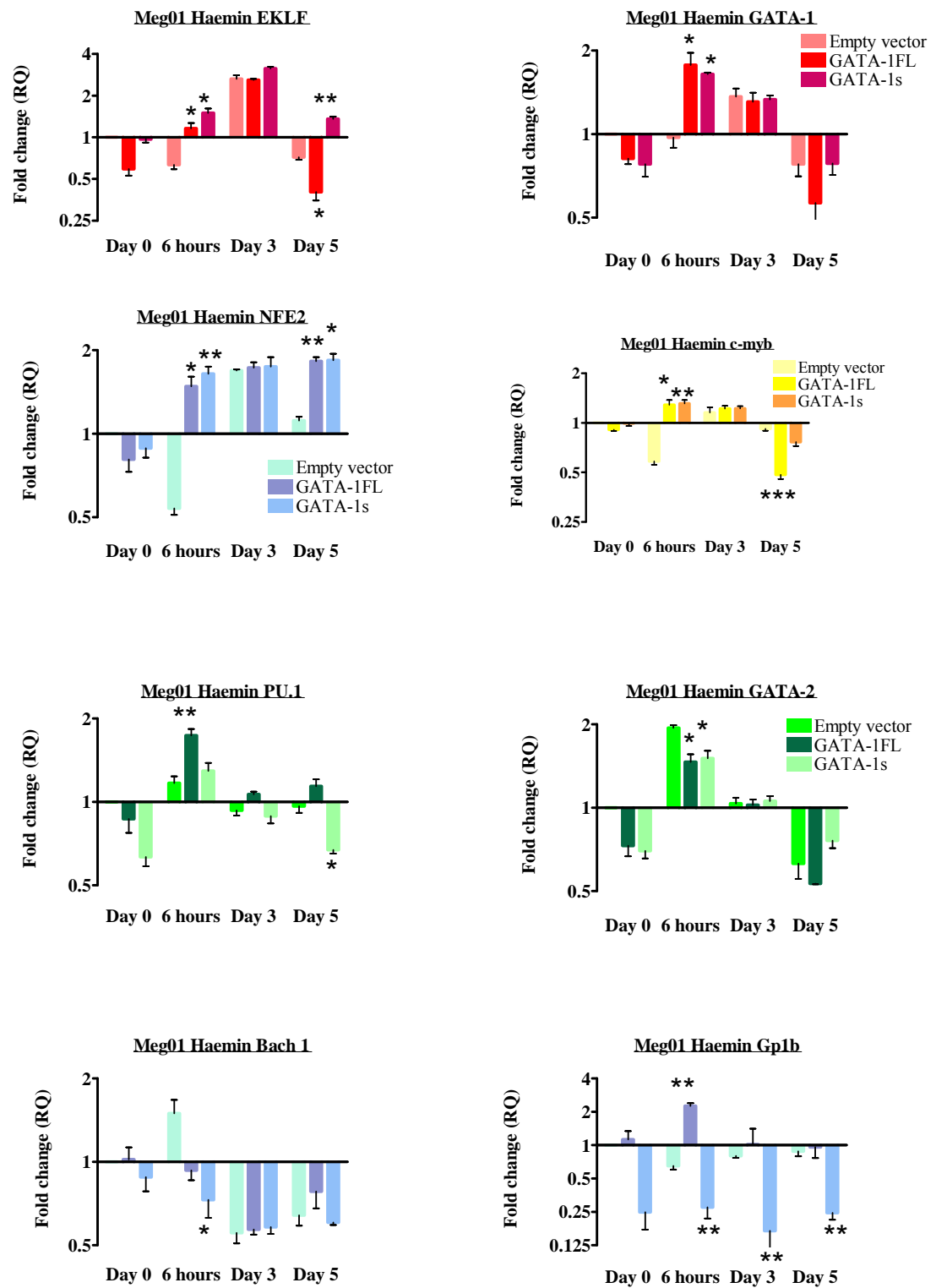
Overall these data suggest that GATA-1FL is driving terminal erythroid differentiation in K562 cells and that GATA-1s expression produces similar gene expression changes but to

a lesser extent. This implies that GATA-1s is acting as a hypomorphic variant of GATA-1FL in these cases. The c-myb result is particularly interesting, it is thought that c-myb needs to be down-regulated prior to terminal differentiation, one of the key findings from the GATA-1s transgenic mice was that fetal liver derived megakaryocytic precursors failed to down-regulate c-myb when differentiated in vitro (Li, *et al* 2005). These results confirm these findings in a completely different experimental system. The same paper also described similar lack of repression with GATA-1s for Ikaros, PU.1, GATA-1 and GATA-2 – also seen above. Whether this reflects lower levels of GATA-1s expression from the transgene, or lack of the N-terminus, remains unclear.

### 5.5.3.3 Gene expression in Meg-01 transgenics

The same experiments were performed in Meg-01 cells. Cells were prepared as described for K562 above. The results of transgene expression are shown in Fig 5.25 overleaf. The majority of genes showed significant differences in expression levels with the 3 constructs although ikaros,  $\beta$ -globin, zeta-globin, ski, Erg3 and cyclin D2 showed similar patterns in all 3. Results for the differentially expressed genes are discussed below:

1. Effect of the GATA-1FL isoform: Significant up-regulation (compared to empty vector) was seen for NFE2 (6 hours and day 5), PU.1 (6 hours), GATA-1 (6 hours), EKLF (6 hours), c-myb (6 hours) and Gp1b (6 hours). Only GATA-2 (6 hours), EKLF (day 5) and c-myb (day 5) were significantly down regulated compared to empty vector.
2. Effect of the GATA-1s isoform: Again the short isoform results mimicked the FL isoform results with upregulation of NFE2 (6 hours and day 5), GATA-1 (6 hours), c-myb (6 hours) and EKLF (6 hours and day 5). Down regulation was seen in GATA-2 (6 hours), Bach-1 (6 hours) and Gp1b (all time points). However on this occasion there was less evidence of GATA-1s acting in a hypomorphic manner with the majority of genes showing equivalent or increased fold change with GATA-1s.



**Fig 5.25 Gene expression changes in Meg-01 transgenics treated with haemin**

\*= $p \leq 0.05$ , \*\*= $p < 0.01$  and \*\*\*= $p < 0.001$

Since GATA-1s is more highly expressed in Meg-01 cells compared to K562 (as evidenced by spontaneous in vitro selection for high GFP expressers and western blot data) this lends support to the theory that at least some of the hypomorphic effects observed in

the K562 assays may be overcome by over-expression of GATA-1s. As discussed in chapter 1 section 1.2.3.3, this has previously been reported with GATA-1s rescue of GATA-1 deficient transgenic mice, where only high levels of transgene expression were able to prevent embryonic lethality (Shimizu, *et al* 2001) . This observation makes the Gp1b results particularly interesting. In this case it would be predicted that the more highly expressed GATA-1s transgene would cause a greater rise in Gp1b than the FL isoform when in fact it produced a significant and sustained down-regulation both with Haemin and TPA treatment. Unfortunately Gp1b is not expressed in K562 cells making it impossible to compare these two cell lines.

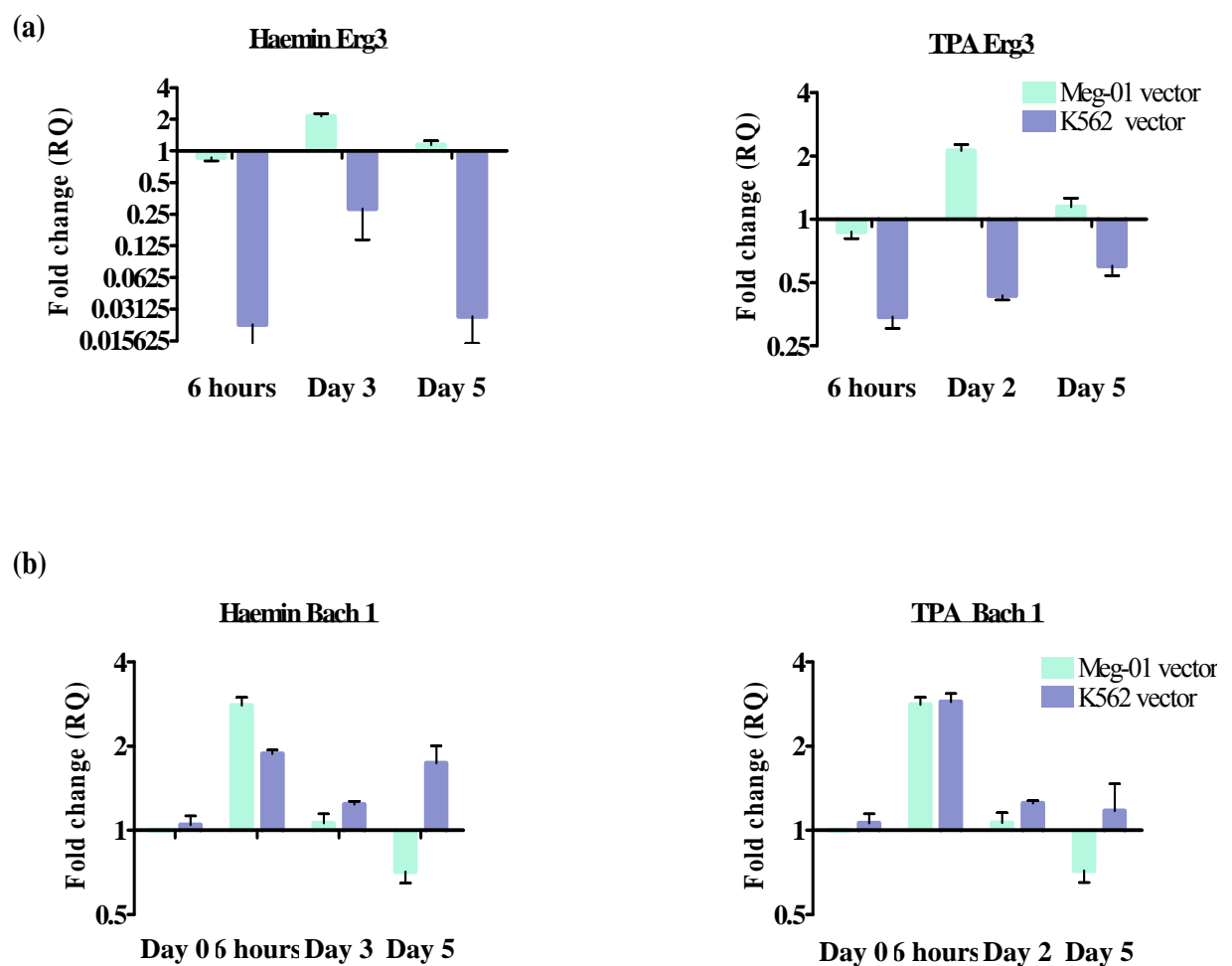
*In summary, the two isoforms show different behaviour with Haemin differentiation in K562 cells. GATA-1FL appears to drive terminal erythroid differentiation with more rapid visible haemoglobinisation of harvested cell pellets. At a gene expression level erythroid genes are transiently upregulated by GATA-1FL and genes involved in alternative lineage choices are down regulated along with c-myb and cyclin D2. The GATA-1s isoform appears to act as a hypomorphic allele with reduced levels of gene activation and repression. This is consistent with previous reports. In Meg-01 cells there is a less marked difference in the expression patterns of the two isoforms, which may reflect higher expression levels of the GATA-1s transgene.*

#### **5.5.4 The impact of GATA-1 isoform expression on Erg3 and Bach-1**

As previously discussed in Chapter 1 section 1.3.2.4 a number of chromosome 21 genes have been proposed as candidates for the exclusive link between trisomy 21, GATA-1s and TAM/AMKL. These include AML-1, Erg3 and Ets2 as possible drivers of a meg-erythroid response and Bach-1 as an inhibitor of terminal differentiation. The Erg3 and Bach-1 genes were included in this gene expression analysis. Results from these analyses did not always produce statistically significant differences between GATA-1FL and GATA-1s transgenics and therefore they are not included in many of the results reported above. However they merit separate consideration because comparison between their expression patterns in K562 (disomic for chromosome 21) and Meg-01 (with trisomy 21) may shed light on their potential roles in DS-AMKL/TAM.

Erg3 is upregulated in a number of AMKL cell lines (Rainis, *et al* 2005). Erg3 is not detectable in K562 cells whilst it is easily detectable in Meg-01 cells compatible with increased expression in the presence of trisomy 21. In addition Erg3 expression levels are

increased on TPA induced differentiation in Meg-01 and Erg3 transcripts become detectable in K562. In order to look at relative expression patterns in K562 and Meg-01 during differentiation transcript levels, from 6 hours onwards, were compared to baseline Meg-01 empty vector cells – the results are shown in Fig 5.26a. As can be seen, baseline expression is higher in Meg-01 and although K562 cells upregulate Erg3 on differentiation they remain several fold lower than levels seen in Meg-01. Interestingly despite its postulated role in megakaryocytic differentiation Erg3 was also upregulated by day 3 of haemin treatment of both cell lines although levels then fell back to baseline. The expression of the various GATA-1 isoforms had no impact on Erg3 expression levels (data not shown) which is unsurprising given that it is not known to be a GATA-1 target gene.



**Fig 5.26 Erg3 and Bach-1 expression levels in K562 and Meg-01 cells**

(a) Response of Erg3 to TPA and Haemin differentiation in Meg-01 and K562 cells, data is shown for the empty vector transgenic cells but there were no significant differences between any of the 3 transgenes (b) Response of Bach-1 to TPA and Haemin differentiation, again empty vector data is shown, the only significant difference between isoforms was seen with GATA-1s during haemin differentiation illustrated in the previous figure (fig 5.25).

Bach-1 has been reported to be upregulated in 2 independent microarray studies of DS-AMKL/TAM (Bourquin, *et al* 2006, Ge, *et al* 2006) and was also shown to be rapidly upregulated on GATA-1 rescue of a GATA-1 deficient cell line (Welch, *et al* 2004). Despite this the TLDA plate data showed no evidence of Bach-1 upregulation at baseline in Meg-01 cells compared to K562. During TPA and Haemin differentiation Bach-1 was upregulated in both cell lines at 6 hours and then down-regulated. In general expression levels were lower in Meg-01 than K562 suggesting that trisomy 21 in these cells did not increase levels during differentiation (as shown in Fig 5.26b). In addition the only difference between the transgenes was increased repression of Bach-1 by GATA-1s during haemin differentiation in Meg-01 (illustrated in Fig 5.25 bottom graph). This suggests that, even in the context of GATA-1s expression, Bach-1 is not over-expressed. Therefore these data do not lend support to the theory that Bach-1 is over-expressed in the context of trisomy 21, although it is possible that this varies with different cellular contexts.

*In conclusion, gene expression data support a potential role for Erg3 over-expression as a driving force for meg-erythroid expansion in trisomy 21. They do not, however, provide any evidence that Bach-1 is upregulated at baseline in trisomy 21 or that the GATA-1s isoform leads to upregulation of Bach-1 expression.*

### **5.5.5 Alternative approaches to studying the effects of isoform expression**

Although the studies above yielded a number of interesting results, one major problem with interpreting the results of isoform over-expression in K562 and Meg-01 cells is the fact that these cells express their own endogenous GATA-1FL and GATA-1s at relatively high levels. This is likely to obscure the effects of the expressed transgenes. This does not necessarily invalidate the results (and in fact may make any observed differences even more likely to be significant) but it considerably complicates the analysis and further investigation of these results. Since GATA-1 is such a fundamental transcription factor for megakaryocytic and erythroid differentiation, it is highly unlikely that any haematopoietic cell line capable of differentiation along these lineages would not show similarly high levels. The use of cells from other lineages is also impossible as GATA-1 requires cell specific co-factors in order to exert its transcriptional effects and therefore any alterations in GATA-1 activity in these cell lines could not be directly attributed to the transgene. This need for cofactors was confirmed by the failure of reporter gene assays in the HEK cell line described in section 5.2.4.3 above.



The approach of gene targeting in murine ES cells is obviously an excellent solution to this problem but, as described in chapter 4, is technically challenging. One other way of addressing this problem would be to knock-down endogenous GATA-1 expression in K562 and Meg-01 cells (using siRNA technology) and then look at the effect of transgenic expression of GATA-1 isoforms in these cells. This is facilitated by the fact that the transgene is murine GATA-1 and the cell lines are of human origin and therefore siRNA duplexes carrying sequence identity to human GATA-1 will knock down endogenous GATA-1 whilst leaving the transgene unaffected. In order to further explore some of the interesting results from the studies above, this approach was attempted as described below:

### ***5.5.6 siRNA mediated knockdown of human GATA-1***

siRNA mediated gene knockdown is a relatively new molecular biology tool. It involves the introduction of small (20-25bp) RNA duplexes with complementary sequences to the desired target sequence. Target recognition is followed by mRNA cleavage and therefore reduction of gene expression. Extremely high rates of gene knock-down have been reported in the literature for some genes although it should be noted that not all genes show these dramatic effects and that commercial siRNAs vary considerably in their efficacy (as discussed below). Many of these commercially available siRNAs are designed using computer algorithms which analyse the target mRNAs and identify theoretically optimal sequences likely to produce good levels of knockdown, the chances of these actually producing in vitro knockdown can be enhanced by targeting more than one section of the mRNA and pooling the individual duplexes before introduction into cells. From this discussion it can be seen that the vast majority of commercially derived siRNAs are not functionally tested prior to despatch.

When devising siRNA experiments it is vital to include appropriate controls. Four controls are generally needed:

**1. A control for non-specific inhibition of gene expression or protein synthesis in the cell:** introduction of RNA duplexes into mammalian cells can activate an interferon response which results in a general inhibition of protein synthesis and eventual cell death. In addition transfection of siRNA does not always produce entirely sequence specific effects. Besides the interferon response, other undesired changes can occur in the cell's gene expression profile, these are usually referred to as off-target effects. In order to control for the interferon response and off-target effects it is necessary to include a negative siRNA control i.e. an siRNA that is non-functional and non-targeting. If this

negative control leads to a reduction in expression of the target gene then induction of off-target effects or the interferon response can be inferred. A commercial fluorescently tagged non-functional siRNA was chosen (siGLO RISC-free siRNA).

**2. A control for non-specific effects of activation of the siRNA processing pathway:** a second control involves the introduction of non-targeting but functional siRNA capable of interaction with RISC and processing within the siRNA pathway (described in chapter 1 section 1.1.3.3.2). A commercial validated control agent (siCONTROL) was used.

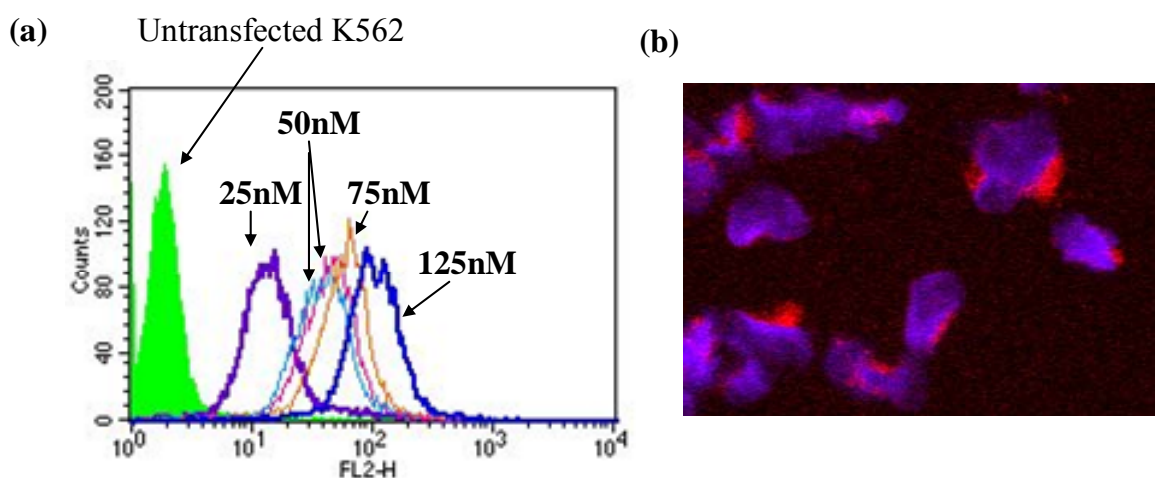
**3. A control for transfection efficiency:** differences in the amount of gene knockdown between cells may reflect differences in transfection rates rather than differences in efficacy of the introduced siRNA – this is particularly important during the optimisation phase where different concentrations of siRNA are being tested. This was achieved by use of the siGLO RISC-free siRNA which acts as a dual control for off-target effects (discussed above) and, by virtue of a fluorescent tag (Rhodamine), for transfection efficiency.

**4. Multiplicity controls to confirm gene specific knockdown:** finally it is useful to be able to perform multiplicity controls. This involves using a pool of siRNAs directed against various regions of the gene. If a specific phenotype is detected on gene knockdown it is then possible to prove that this is a result of specific gene knockdown rather than off-target effects by showing that each individual siRNA (components of the pool) is capable of reproducing the observed phenotype.

Bearing all these factors in mind human GATA-1 gene silencing in K562 cells (wild-type) was attempted, with a view to proceeding to gene silencing in transgenic K562 cells and Meg-01 (leaving transgenic murine GATA-1 isoform expression intact) following proof of principle in K562. Unfortunately, at the time of devising these experiments there was no published data on human GATA-1 siRNA mediated knockdown and therefore no validated siRNA to choose. On the basis of a good reputation in the field and the availability of pools of siRNA duplexes it was decided to use a human GATA-1 siRNA pool (SMARTpool) designed by Dharmacon Inc. The sequences of the 4 pooled duplexes are listed in table 2.12, chapter 2. It is notable that one of these duplexes (duplex 6) binds to a target sequence in exon 2 of the GATA-1 mRNA. In theory this duplex would selectively knock-down the GATA-1FL isoform, as the alternatively spliced GATA-1s transcript lacks exon 2 and therefore would escape cleavage.

### 5.5.6.1 Optimisation of transfection efficiency

siRNA can be introduced to cells using lipid-mediated transfection or electroporation. Exposure to transfection reagents can affect gene transcription so a mock-transfected control was run for all experiments. A well established siRNA transfection agent was tested – Lipofectamine 2000 (Invitrogen). To examine transfection efficiency the non-targeting fluorescently tagged –siGLO RISC-free siRNA was used. Cells were cultured and transfected as described in Materials and Methods section 2.13.7. 4 different concentrations of siGLO RISC-free siRNA (25-125nM) and 3 different quantities of lipofectamine 2000 (2.5µl-6µl) were tested. As shown in Fig 5.27a all combinations gave excellent transfection efficiencies. To ensure that these FACS results were not due to adherence of transfection complexes to the surface of transfected cells rather than entry into the cells, confocal microscopy (as described in materials and methods, section 2.13.6) was also performed with nuclei counterstaining with DAPI. 3D visualisation was achieved using Z-stack software and confirmed distribution throughout the cytoplasm. Fig 5.27b shows a representative confocal image. Following this optimisation experiment 5µl of lipofectamine and 50nM of siRNA was used for subsequent experiments.

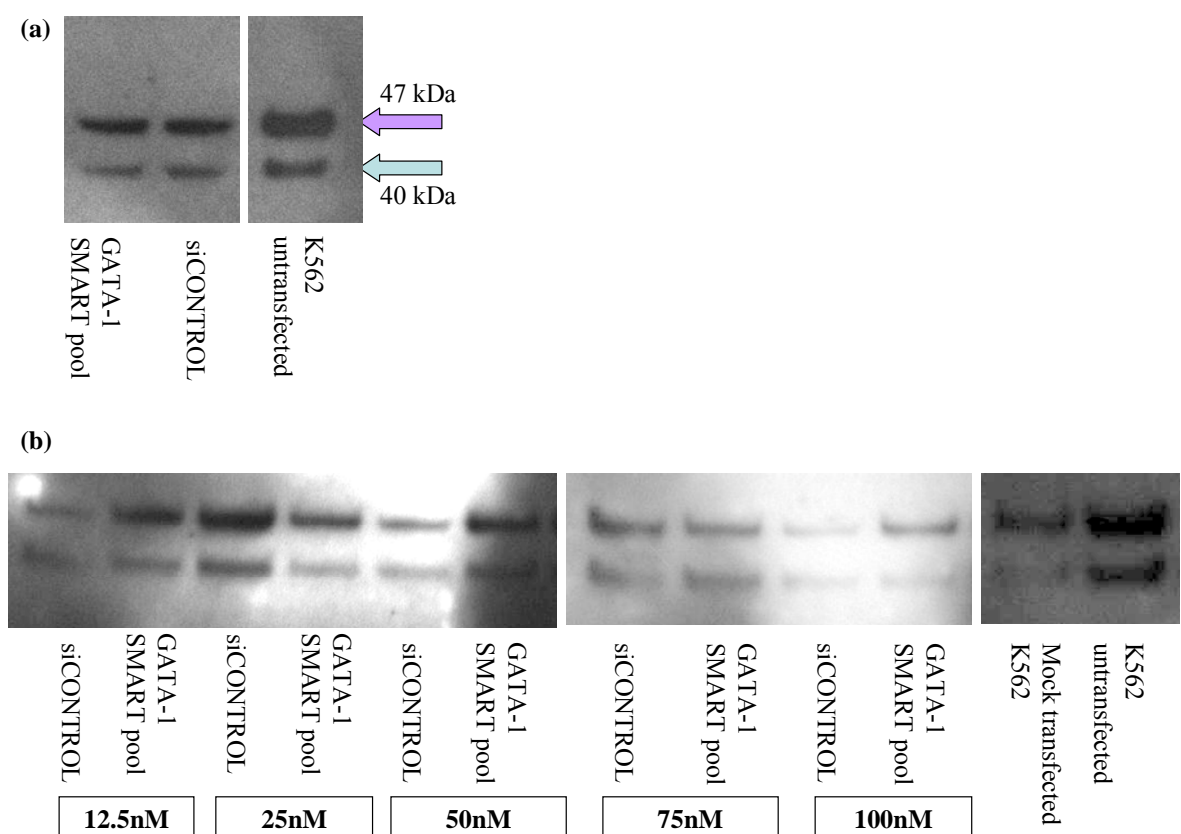


**Fig 5.27 Estimation of siRNA transfection efficiency using Rhodamine labelled RISC free siRNA**

(a) FACS analysis of transfected cells looking at Rhodamine fluorescence (FL2 channel), various concentrations of siRNA from 25nM to 125nM were tested, all gave acceptable transfection efficiencies with evidence of a dose response. 5µl of lipofectamine was used for all transfections except one (50nM pale blue line) which used 2.5µl, this did not dramatically affect transfection efficiency (b) representative confocal image, nuclei are counterstained with DAPI (blue) and Rhodamine fluoresces red. The intra-cytoplasmic location of the red signal was confirmed by Z-stack software (images not shown).

### 5.5.6.2 Effects of siRNA mediated GATA-1 targeting

The next step was to attempt GATA-1 specific gene knockdown. 25nM of siGLO RISC-free siRNA and 25nM of either siCONTROL or SMART pool GATA-1 siRNA were introduced into K562 cells. Mock transfected and untransfected cells were also used. After 48 hours cells were harvested and counted. Cells ( $4 \times 10^5$ ) were used in a standard nuclear extract protocol (section 2.9.2 chapter 2), the remainder of the cell pellet was resuspended in FACS buffer and analysed for transfection efficiency. Excellent transfection efficiencies were again seen with a log shift in FL2 channel fluorescence in both SMART pool and siCONTROL transfected cells (data not shown). Western blotting was performed (as described in section 2.9.4) and equivalent protein loading was confirmed by Ponceau S staining of the membrane (section 2.9.5). Results are shown in fig 5.28a. It can be seen that both the SMART pool and siCONTROL reagents produced slightly lower levels of GATA-1 protein than untransfected K562 cells but that there was no evidence of specific GATA-1 silencing by the SMART pool.



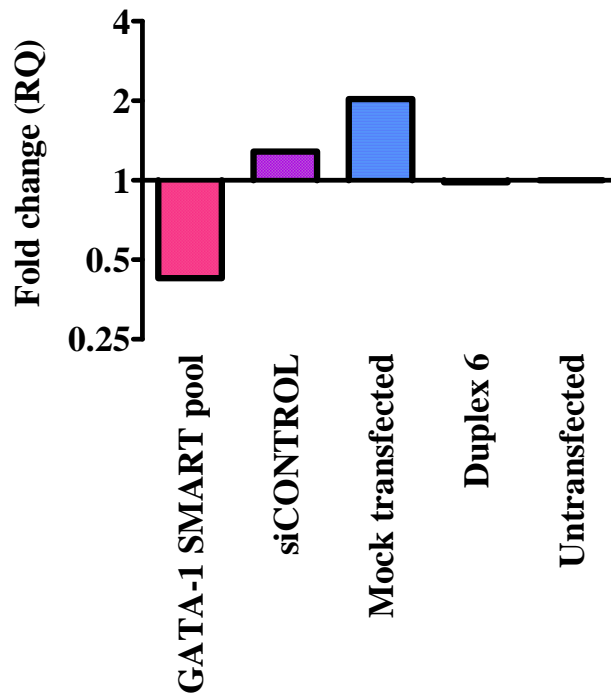
**Fig 5.28 siRNA knockdown of GATA-1 at the protein level**

(a) western blotting reveals no evidence of protein knockdown 48 hours post transfection with siRNA (b) further attempts at knockdown with nuclear extracts at 72 hours post transfection using a variety of concentrations of siRNA (shown in boxes beneath figure) still failed to show significant knockdown at the protein level when control non-targeting siRNA (siCONTROL) was compared to GATA-1 specific siRNA (GATA-1 SMART pool)

In case this effect was due to inadequate levels of siRNA in the cells the experiment was repeated using a variety of concentrations of SMART pool or siCONTROL siRNA from 12.5nM to 100nM. Cells were left 72 hours prior to harvesting. Again, despite excellent transfection efficiencies (data not shown), and good cell viability (>90% with all transfectants, as measured by Trypan blue exclusion), there was no evidence of GATA-1 specific silencing as shown in fig 5.28b. In fact, the lanes with the apparently lower GATA-1 protein levels all represented siCONTROL transfected cells and these differences were accounted for by small variations in the protein loading per lane (as seen by Ponceau S staining).

As a final attempt to show specific knockdown of GATA-1 by the SMART pool it was decided to look for alterations in mRNA levels. Since mRNA represents the primary target of the siRNA, knockdown of mRNA may be more rapid and complete than knockdown at the protein level. 25nM of siGLO RISC-free siRNA and 75nM of SMART pool, duplex 6 (no targeting of GATA-1s) or siCONTROL siRNA were transfected into K562 cells, a mock transfection control was also set up. Cells were harvested at 48 hours and following RNA extraction (RNeasy kit, section 2.8.1.2) and cDNA synthesis from 2µg starting RNA (Affinity Script II, section 2.8.4) the cDNA template was diluted 1 in 10 and qPCR performed. Custom GATA-1s primer and probes (table 2.11, chapter 2) were used to detect GATA-1s transcripts and levels were compared to the 18SrRNA housekeeping gene, K562 untransfected cells were used as the calibrator. Results are shown in fig 5.29. In contrast to the western blotting results, a two-fold reduction in GATA-1s mRNA levels was seen with the SMART pool compared to either untransfected controls or siCONTROL transfected controls. The use of duplex 6 provided an additional negative control as this duplex, although showing sequence specificity for the GATA-1 target mRNA, should not alter GATA-1s mRNA levels (an exon 1-3 spliced transcript) as it binds to exon 2. This was confirmed with no change in gene expression levels compared to untransfected controls. Therefore, it appears that the GATA-1 SMART pool does achieve some gene specific knock-down at the mRNA level but that this is not translated into reduced protein levels at 48 or 72 hours.

### siRNA changes in GATA-1s gene expression levels



**Fig 5.29 Results of qPCR analysis of siRNA mediated GATA-1 knockdown**

#### 5.5.6.3 Discussion of siRNA results

Despite demonstration of specific GATA-1 knockdown at the mRNA level at 48 hours, GATA-1 protein knockdown could not be detected at 48 or 72 hours. To be of practical use in assessment of the impact of isoform expression on GATA-1 function, endogenous GATA-1 knockdown would have to occur at the protein level and be sustained and fairly complete. Therefore this technique was not pursued further. Various possibilities for lack of GATA-1 knockdown at the protein level include:

**1. A long half life of the GATA-1 protein:** although the GATA-1 half-life has not been formally documented in the literature, the demonstration by other investigators of undetectable GATA-1 protein at 48 hours following successful siRNA knockdown (Li, *et al* 2008a) (discussed further below) suggests the 48 hour time point should have been sufficient to detect reductions in protein content.

**2. Incomplete mRNA knockdown:** qPCR results show an approximately 2 fold reduction in mRNA levels. This may be insufficient to cause a drop in protein levels particularly if GATA-1 is subject to additional post-transcriptional or post-translational control

**3. Autoregulation:** Since GATA-1 is known to autoregulate its expression, a transient fall in GATA-1 protein may be rapidly compensated for by increased gene transcription.

**4. A Growth/survival disadvantage in knockdown cells:** since GATA-1 is known to promote survival and prevent apoptosis of erythroid precursors it is possible that cells with reduced levels of GATA-1 expression have a growth or survival disadvantage in culture. Therefore on more prolonged culture there would be a selective pressure for cells that had not undergone efficient gene knockdown.

Subsequent to these studies three attempts at GATA-1 gene targeting by siRNA have been reported. Using a murine Dharmacon siRNA (D-045656-04) one group reported knockdown of murine GATA-1 although there was still easily detectable GATA-1 protein by Western blot and GATA-1 induced luciferase activity was about 70-80% of that seen with a control siRNA. Therefore this knockdown was modest at best (Masuda, *et al* 2007). In a human cell line (HT93) GATA-1 induction by ATRA treatment could be reduced at the protein level by transfection of a commercially available human GATA-1 siRNA (#sc-29330, Santa-Cruz Biotechnology) at 10nM although again protein was still easily detectable by western blotting (Yoshida, *et al* 2007). Interestingly the group tried a variety of siRNA concentrations from 10-100nM and showed maximal specific inhibition at the lowest concentration. Finally one group have recently reported successful use of a commercial SiRNA (GenePharma Biotechnology) at a concentration of 1µM in K562 cells which showed moderate GATA-1 knockdown at 24 hours and complete absence of detectable expression at 48 hours by western blotting, although the control siRNA also showed moderate reduction in GATA-1 levels at this time point as measured by western blot. They also reported functional reduction in GATA-1 in that a GATA-1 reporter luciferase assay showed a reduction in transactivation from 100% to 50% with GATA-1 siRNA knockdown (Li, *et al* 2008a). These published results suggest GATA-1 gene knockdown is possible but, unsurprisingly, the level of knockdown varies depending on the siRNA construct used. In addition none of the studies was able to show complete knockdown of GATA-1, particularly at a functional level. This does not necessarily matter when the aim of the experiment is to assess the effects of reducing expression levels and then restoring expression levels on a defined output such as expression of a GATA-1 target gene. However, it would still complicate analysis of GATA-1 isoforms as differences

between the constructs would be difficult to disentangle from subtle differences in the level or duration of (incomplete) endogenous GATA-1 knockdown. Successful use of siRNA for these purposes probably requires advances in siRNA technology allowing more prolonged and effective siRNA delivery with reduced non-specific toxic effects on cell protein synthesis. Given the rate of advance in this field these advances are likely to occur in the near future.

## **5.6 Summary of the results of functional assessment of GATA-1 isoforms in murine and human cell lines**

This chapter has described experiments to investigate whether the two GATA-1 isoforms behave differently during in vitro assays of hematopoietic function. Using a variety of different techniques consistent differences between the two isoforms emerge i.e.

1. The GATA-1FL isoform appears to drive terminal erythroid or megakaryocytic differentiation depending upon the context in which it is expressed. This is seen in murine and human cell lines. In murine ES cells there is an increase in erythroid and CFU-GM colony output compared to wild-type cells. In Meg-01 and K562 cells expression of the GATA-1FL transgene is associated with increased growth rates (particularly in Meg-01) and significant modulation of erythroid and alternative lineage genes at the mRNA level compared to empty vector and GATA-1s expressing cells. On induced differentiation with TPA, GATA-1FL expressing cells show increased DNA ploidy and rapid down regulation of genes required to allow terminal differentiation (*c-myb*, *ski*). On induced differentiation with haemin GATA-1FL rapidly become visibly haemoglobinised (< 24hours) and upregulate erythroid genes.

2. The GATA-1s isoform is much less effective at supporting cellular differentiation. In murine cells expression of GATA-1s leads to almost complete ablation of erythroid and CFU-GM colony output. In human cell lines GATA-1s appears to act in most cases as a hypomorphic allele of GATA-1FL with similar changes in gene expression but to a much lesser extent. In general GATA-1s expression was associated with partial but incomplete differentiation with a reduced ploidy response to TPA treatment and inadequate repression of genes such as *ski* and *c-myb* which require down-regulation prior to terminal differentiation. In addition, expression of the GATA-1s isoform appeared to favour megakaryocytic over erythroid patterns of gene expression and features such as the upregulation of *ski* suggest it may promote maintenance of a bi-potential meg-erythroid



precursor. The inability to up-regulate Gp1b expression, even when GATA-1s is expressed at high levels, suggests that the GATA-1s does not drive terminal megakaryocytic differentiation.

Two human cell lines, K562 and Meg-01, were studied to test the hypothesis that the GATA-1s isoform may only manifest some biological properties in the presence of trisomy 21. The comparison between these two cell lines is complicated because of differences in their stage of hematopoietic commitment, baseline gene expression profiles and the fact that they spontaneously segregated themselves into opposite high and low expressing transgene populations in culture (i.e. high expression of GATA-1FL and low expression of GATA-1s in K562 and vice versa in Meg-01). This observation is intriguing and cannot easily be explained by observed differences in growth rates or gene expression. However, it is extremely interesting that the cell line carrying trisomy 21 (Meg-01) spontaneously selected GATA-1s whilst the non-trisomy 21 cell line selected GATA-1FL. This mirrors the clinical observation that GATA-1s associated malignancy/myeloproliferation is only ever seen in the context of trisomy 21 leading to the hypothesis that the combination of GATA-1s and trisomy 21 confers a unique selective advantage not seen in disomic cells. In addition gene expression data suggest that *Erg3* may be over-expressed in the presence of trisomy 21 but that *Bach-1*, another chromosome 21 candidate gene, does not show evidence of over-expression in this model system. Possible explanations for these phenomena are put forward in the discussion, chapter 7.

Finally, these studies using CML cell lines suggest that modulations in GATA-1 levels of expression and isoform production can affect the ability of CML cells to proliferate and terminally differentiate. Taken alongside the discovery of a potentially hyper-functioning GATA-1 mRNA variant in a patient with CML – BC009797 discussed in chapter 3- it was decided to investigate the association between CML and GATA-1 further and in particular to test the hypothesis that GATA-1 mutations may be responsible for progression to blast crisis in CML. These studies are described in the next chapter.

## 6 Analysis of GATA-1 Mutational Status in Chronic Myeloid Leukaemia

---

### 6.1 Background

The molecular basis for progression of chronic myeloid leukaemia from chronic phase to accelerated phase and blast crisis is not well understood. According to the type I and type II mutation hypothesis (Gilliland 2001) the bcr-abl oncoprotein, with aberrant tyrosine kinase activity, would provide a type I mutation and therefore a type II mutation in a gene involved in cellular differentiation would be predicted to be responsible for disease progression. This has been confirmed in a small number of cases (Zhang 2008), but in the majority of cases progression remains unexplained at a molecular level (Melo and Barnes 2007).

As discussed in Chapters 1 and 3, two pieces of circumstantial evidence suggest a possible role for aberrant GATA-1 expression in chronic myeloid leukaemia (CML) i.e.

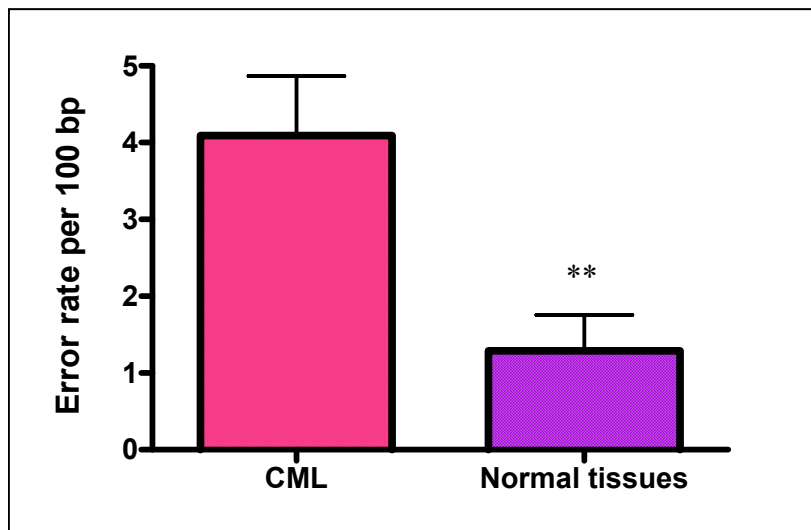
1. The cloning of an aberrant mRNA transcript – BC009797 from the bone marrow of a patient with chronic myeloid leukaemia.
2. The strong morphological resemblance between transient abnormal myelopoiesis and the chronic phase of CML in particular the marked eosinophilia, thrombocytosis and basophilia in both conditions.

Importantly, the study reporting the tight association between GATA-1 mutations and DS-AMKL (Wechsler, *et al* 2002) did not investigate GATA-1 mutational status in CML.

Given the plausible association between perturbations in GATA-1 function and progression of CML, it was hypothesised that mutations in GATA-1 could be involved in the transition from chronic phase to blast crisis in CML. To test this hypothesis, *in silico* analysis was performed followed by GATA-1 mutational analysis in CML patient samples.

## 6.2 In silico analysis

To assess whether aberrant GATA-1 mRNA transcripts are a common feature of CML an in silico analysis was performed looking for all published ESTs aligning to the known GATA-1 mRNA sequence NM\_002049. This search produced 14 ESTs longer than 500bp and a further 14 shorter sequences all derived from fetal tissues (and none from CML). Four of the fourteen longer sequences were derived from patients with CML the others came from normal adults (n=6) or fetal samples (n=4). Comparing the number of mismatches per 100 basepairs in the longer sequence set (so avoiding bias towards sequencing anomalies at the beginning of short transcripts) a statistically significant difference between CML samples and normal individuals was seen ( $p=0.007$ , fig 6.1). This finding needs to be interpreted with caution as ESTs commonly contain sequencing errors but it adds weight to the hypothesis that CML may be associated with GATA-1 mutations.



**Fig 6.1 EST sequence error rate in CML and normal individuals**

Comparison of mismatches per 100bp in EST sequences derived from patients with CML (n=4) and normal individuals (n=10),  $p=0.007$  using two-tailed t-test.

## 6.3 Mutational analysis using patient samples

### 6.3.1 Samples

Genomic DNA from 24 patients with CML in accelerated phase (n=1), myeloid blast crisis (n=15) or lymphoid blast crisis (n=8) was kindly provided by Dr Gordon Strathee

(Beatson Laboratories, Glasgow). This DNA was extracted from either peripheral blood or bone marrow samples. Leukocytes were isolated following ammonium chloride red cell lysis and CD34 selection was performed in some cases. DNA concentration varied from 120ng/μl to 370ng/μl. A human embryonic kidney cell line – HEK293 was used to provide control DNA, extracted using cell lysis buffer and ethanol precipitation as described in materials and methods (section 2.7.1). Methods to optimise mutation detection and results of this analysis are described below.

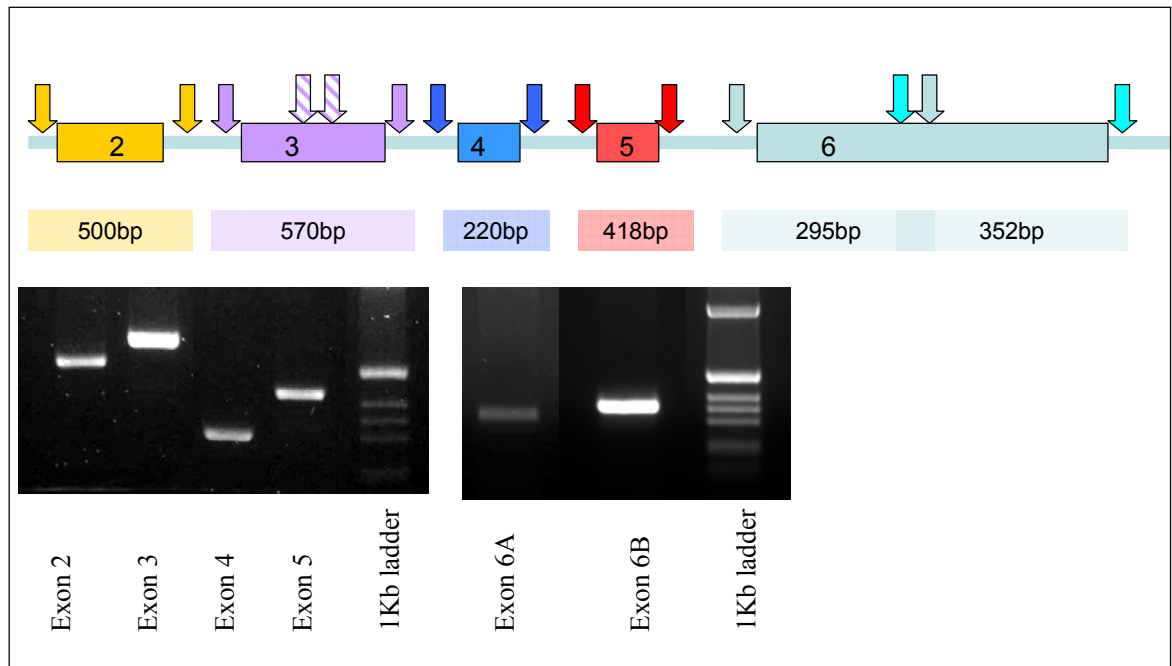
### ***6.3.2 Optimisation of methods to detect GATA-1 mutations***

The in silico analysis above showed mismatches in CML samples scattered throughout all 5 coding exons. Therefore it was decided to assess the entire coding sequence of GATA-1 for mutations. A number of possible methods exist for analysis of mutations in DNA (Cotton 1997). These include single-strand conformation polymorphism (SSCP), Denaturing gradient gel electrophoresis (DGGE), mismatch cleavage methods, heteroduplex analysis including the use of denaturing high-performance liquid chromatography (DHPLC) and finally direct DNA sequencing (Cotton 1997). All except the latter rely on identification of a mismatch between wild-type and mutant sequences, the location and chemical nature of this mismatch still needs to be identified by sequencing. Therefore these methods are useful for screening multiple samples followed by targeted sequencing of only the abnormal products. Having assessed the various methods and considering the relatively small number of samples involved in this study, it was decided to test the sensitivity and specificity of direct sequencing. Experiments were initially performed on genomic DNA from HEK293 to establish methods before moving on to analysis of patient samples.

### ***6.3.3 Approach 1: PCR of coding exons followed by direct sequencing of products***

Amplification of DNA fragments for direct sequencing was performed by designing PCR primers to anneal to introns approximately 50-100 basepairs upstream and downstream of coding exons 2-6 to produce products of 200-570bp in length. Exon 6 is 518bp long and so was divided into two smaller fragments 6.1 and 6.2. The position of PCR primers, predicted products and representative electrophoresis gels are shown in fig 6.2, primer sequences are listed in materials and methods 2.4.5. These initial experiments confirmed that the primers produced a single band on gel electrophoresis. The PCR products were

purified (QIAquick PCR purification kit, materials and methods section 2.7.5) and sent for standard Sanger sequencing (Agowa.de). Results confirmed that the PCR products matched the expected GATA-1 exons (data not shown).

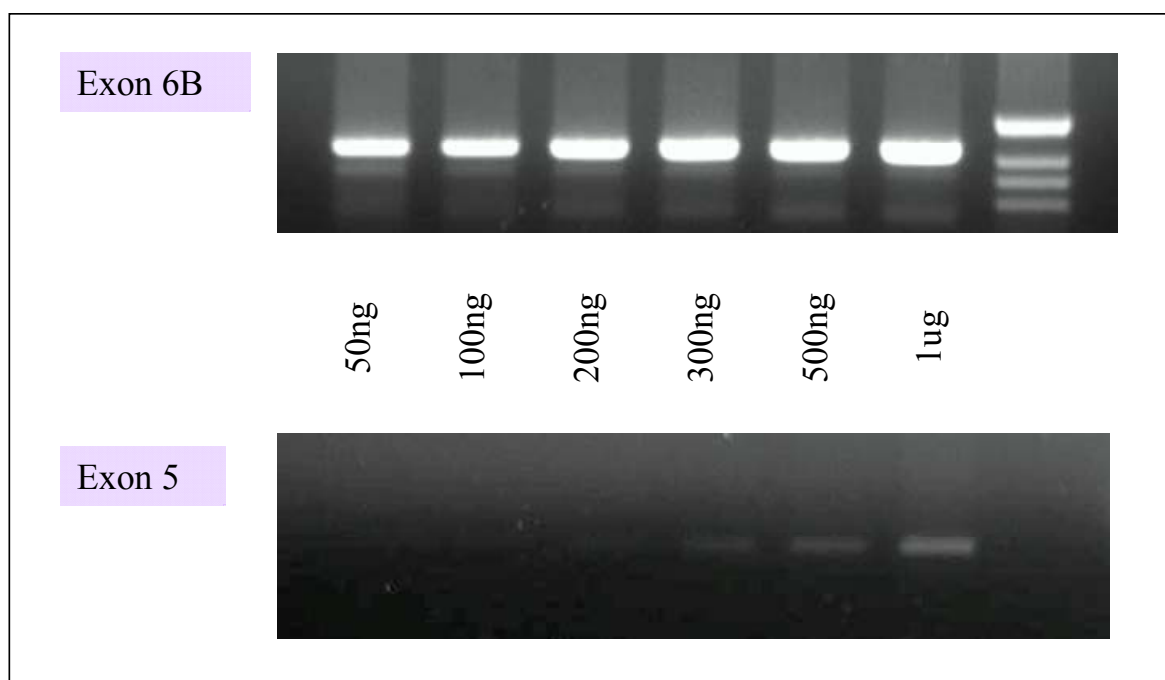


**Fig 6.2 PCR of coding exons with direct sequencing of products**

The position of primers is indicated with filled arrows. The striped arrows in exon 3 represent additional splitting of this exon used approach 2, otherwise primer positions are approximately the same with both approaches. PCR amplification produces a single band of the correct size on gel electrophoresis (2% agarose gel). The identity of this band was confirmed on sequencing.

Since the amount of template DNA was limited it was important to test whether accurate sequencing could be achieved with low amounts of starting template. Therefore, a titration analysis was performed. The primer sets amplified DNA with variable efficiency so template titration was performed using a low efficiency primer set (exon 5) and a high efficiency combination (exon 6B). The results are shown in Fig 6.3 which shows that 50ng of template is sufficient for a strong product band using exon 6B primers but that 200ng is probably required to achieve adequate amplification with the less efficient exon 5 primers. In order to test whether direct sequencing could still produce good quality data from these small amounts of template 150ng of HEK 293 genomic DNA was used in the PCR reaction. Following amplification using a Taq polymerase (ReddyMix, materials and methods section 2.10.1) fragments were purified using a QIAquick PCR purification kit. DNA concentration was estimated by gel electrophoresis and estimation of absorbance at 260nm. Approximately 100-200ng of DNA was sent for commercial sequencing (Agowa.de). Good data were obtained from analysis of exon 3, 5 and 6A amplification products but other samples 2, 4 and 6B showed high background or were unable to deliver

data. Although it would certainly have been possible to optimise this method further by the design of alternative primers, alteration of PCR reaction conditions, or use of alternative DNA polymerases with higher amplification efficiency, it was decided to explore other options in parallel as discussed below.



**Fig 6.3 Titration of genomic DNA template for high and low efficiency PCR reactions**

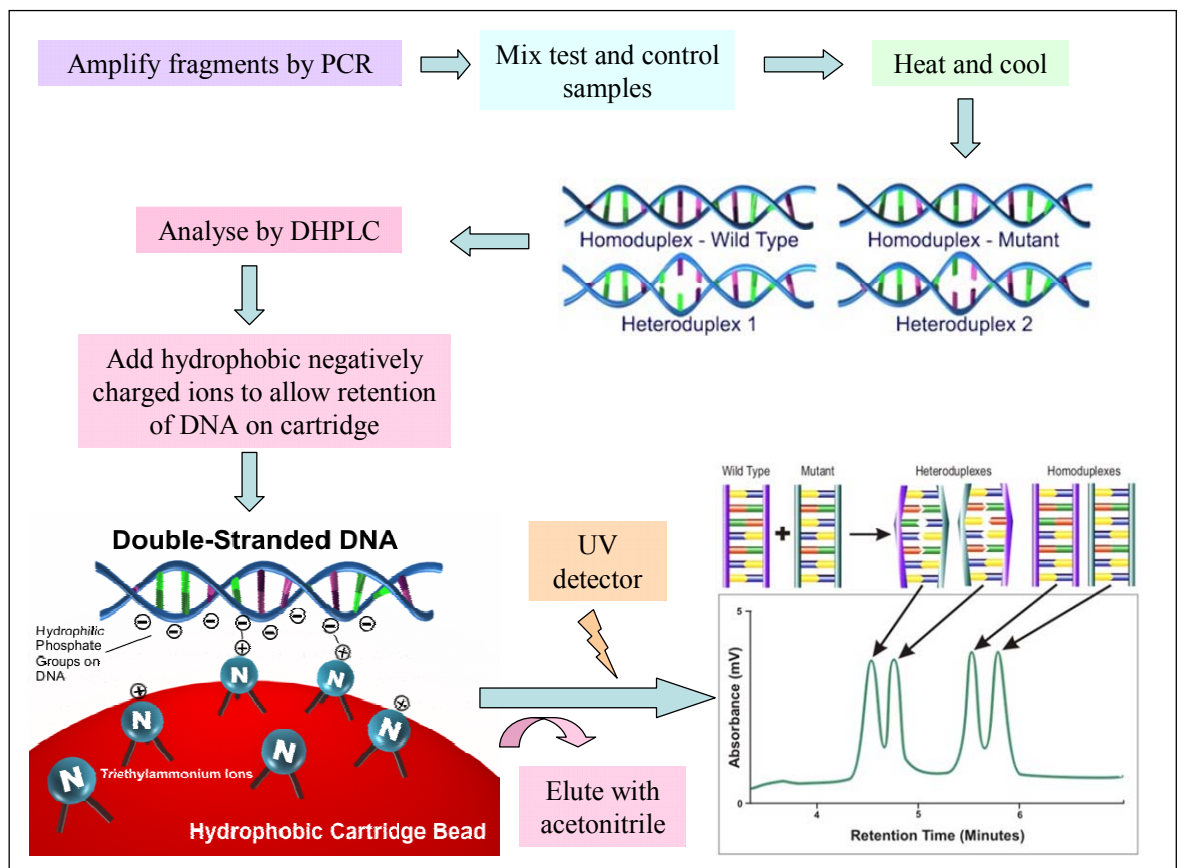
HEK293 genomic DNA was diluted to produce stock solutions of 50ng/ $\mu$ l, 100ng/ $\mu$ l, 300ng/ $\mu$ l and 500ng/ $\mu$ l. 1-2 $\mu$ l of stock solution was added to a standard PCR reaction (ReddyMix plus 1.25 $\mu$ l of forward and reverse primers) along with water to make a final template volume of 2.5 $\mu$ l. 35 cycles of PCR were performed using optimised conditions for each primer pair.

#### **6.3.4 Approach 2: PCR of coding exons followed by analysis of products by DHPLC**

##### **6.3.4.1 Principles of WAVE technique**

A second approach was to screen for mutations using detection of heteroduplexes by DHPLC followed by sequencing of any samples showing abnormal chromatograms (Xiao and Oefner 2001). A protocol was obtained from collaborators (Dr P Vyas and Dr C Fisher) who routinely perform GATA-1 mutational analysis as part of a neonatal screening programme for children with trisomy 21 (Vyas and Roberts 2006). Heteroduplex analysis can be performed in a variety of ways including by exploitation of different mobilities on non-denaturing gels but use of DHPLC allows automation and standardisation, which is particularly important for clinical applications (Cotton 1997). In this analysis a commercial

DHPLC system optimised for heteroduplex analysis was used (WAVE – Transgenomics, Omaha, USA). An outline of the principles behind this technique is shown in Fig 6.4.



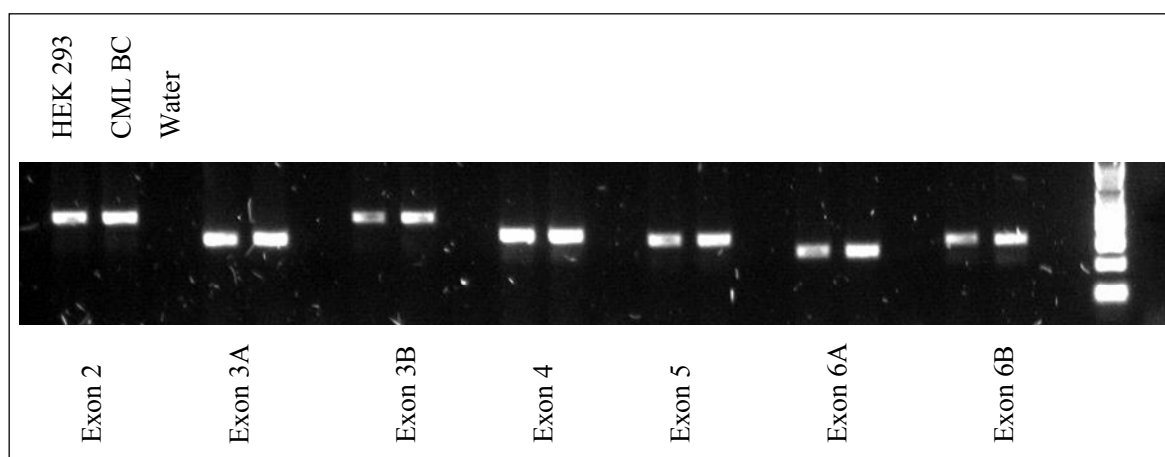
**Fig 6.4 Principles of WAVE mutation screening using DHPLC**

Essentially this involves amplification of test and control (ie non-mutated) DNA fragments by PCR. These reaction products are mixed and then heated to separate the DNA strands followed by cooling to allow reannealing. On cooling up to 4 possible duplexes are produced – control-control, test-test, control-test and test-control (Fig 6.4). These products are then mixed with the buffer triethylammonium acetate (TEAA). In solution this compound produces positively charged triethylammonium (TEA) ions with hydrophobic and hydrophilic ends. The positive charge on these ions associates with the negatively charged phosphate backbone of the DNA duplexes creating a hydrophobic “coating” on the DNA. This solution then passes through a cartridge (DNASep –Transgenomics) containing hydrophobic beads which retain the hydrophobic DNA-TEA complex. The cartridge is then heated to cause partial denaturation of the DNA duplexes and an increasing gradient of the organic solvent acetonitrile (ACN) flows through the cartridge. ACN provides a small hydrophilic counter-ion to cause disruption of the TEA-DNA bridge. This leads to elution of DNA complexes from the hydrophobic beads with mismatched complexes eluting first and then homoduplexes. The DNA is detected by passage through a UV

detector measuring  $A_{260}$  and this measurement is plotted against time. If no mutation exists then the DNA will elute as a single homoduplex peak (sometimes with a prominent notch). If a mutation is present then between 1 and 2 additional peaks are produced (Fig 6.4). Each test DNA fragment will have an optimum elution temperature which is predictable by melt profile software.

### 6.3.4.2 WAVE analysis of CML patient samples

The design of PCR primers and optimisation of elution temperature conditions had already been performed for all 5 coding exons of human GATA-1 by our collaborators. The primer locations (but not sequences) essentially mirrored those used for direct sequence analysis above (as shown in Fig 6.2) with the exception of exon 3 which is 377bp long and was split into two fragments for amplification and analysis. All primer sequences are listed in section 2.4.5 materials and methods, along with the optimised elution temperature for each fragment. Primers were tested on HEK 293 genomic DNA and a CML blast crisis sample (M1277BC) and bands of the correct size were produced in all cases shown in Fig 6.5.



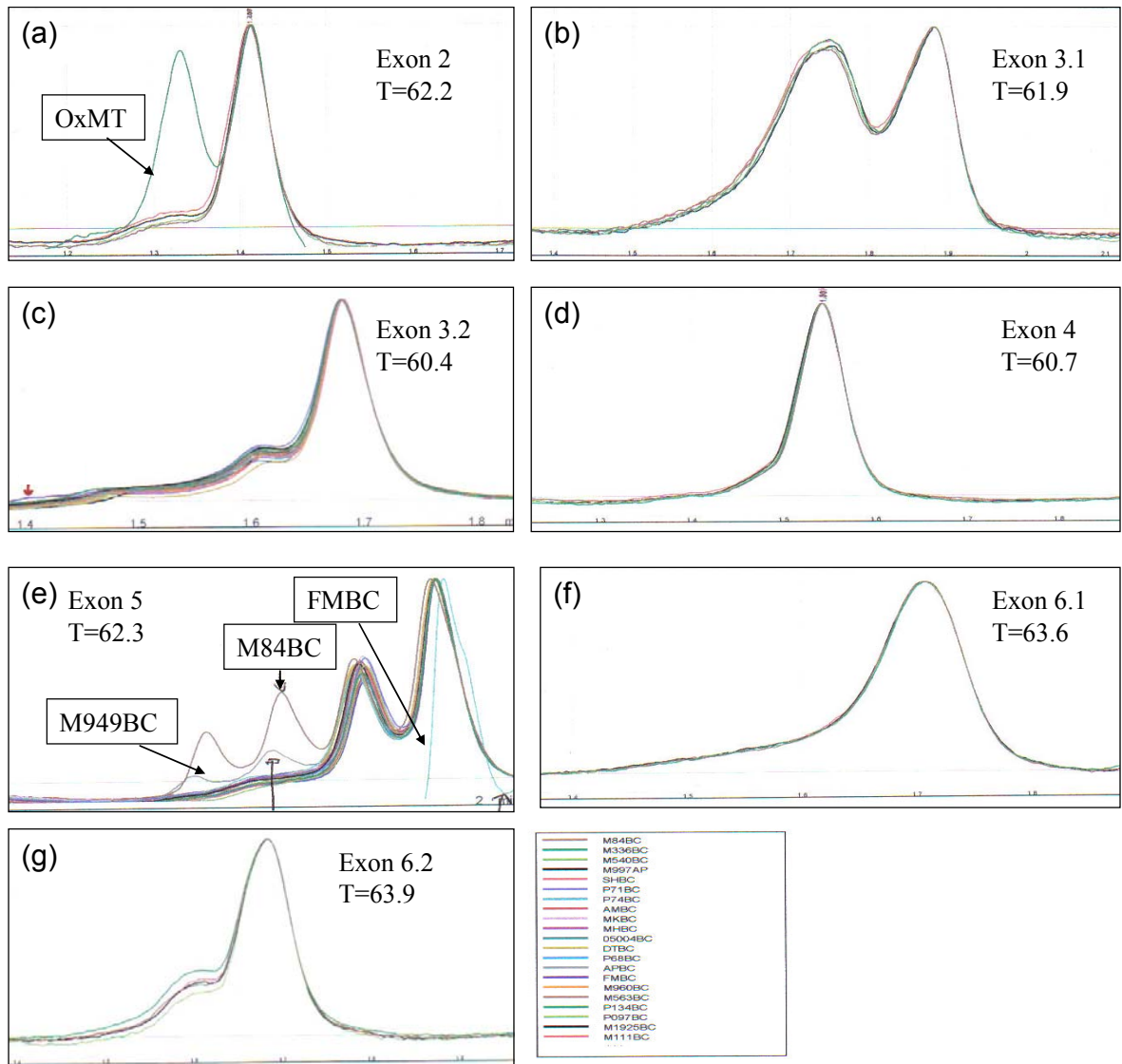
**Fig 6.5 Amplification of coding exons of GATA-1 for analysis by DHPLC**

**PCR was performed as described in section 2.11, 5µl was run on a 2% agarose gel to confirm amplification as shown. The residual sample was then used for DHPLC analysis (WAVE) +/- direct sequencing.**

DNA fragments were amplified using AmpliTaq Gold DNA polymerase (Applied Biosystems) mixed with a proof reading DNA polymerase (Pwo, Roche). A hot-start protocol was used to increase specificity and yield (as detailed in section 2.11). Amplification was confirmed by running 5µl of product on a 2% agarose gel (fig 6.5). 10µl of sample was then mixed with 10µl of control DNA fragments amplified using the same primers from HEK293. Heteroduplexing was performed as described in section 2.11.2. The samples were then run on the WAVE platform. Each DNA fragment was analysed



separately with two different elution temperatures (determined by melt profile software). A known mutant sample with a single basepair mismatch in exon 2 (OxMT) was run as a positive control. Representative chromatograms are shown in Fig 6.6. Overall 2 samples showed abnormal chromatograms: M949BC (myeloid BC) and M84BC (myeloid BC), both in the exon 5 fragment (Fig 6.6e and Fig 6.7), in addition to the exon 2 mutation in the positive control (Fig 6.6a). Another sample FMBC (myeloid BC) showed an aberrant chromatogram (Fig 6.6e), although the trace did not correspond to normal patterns of heteroduplex elution.

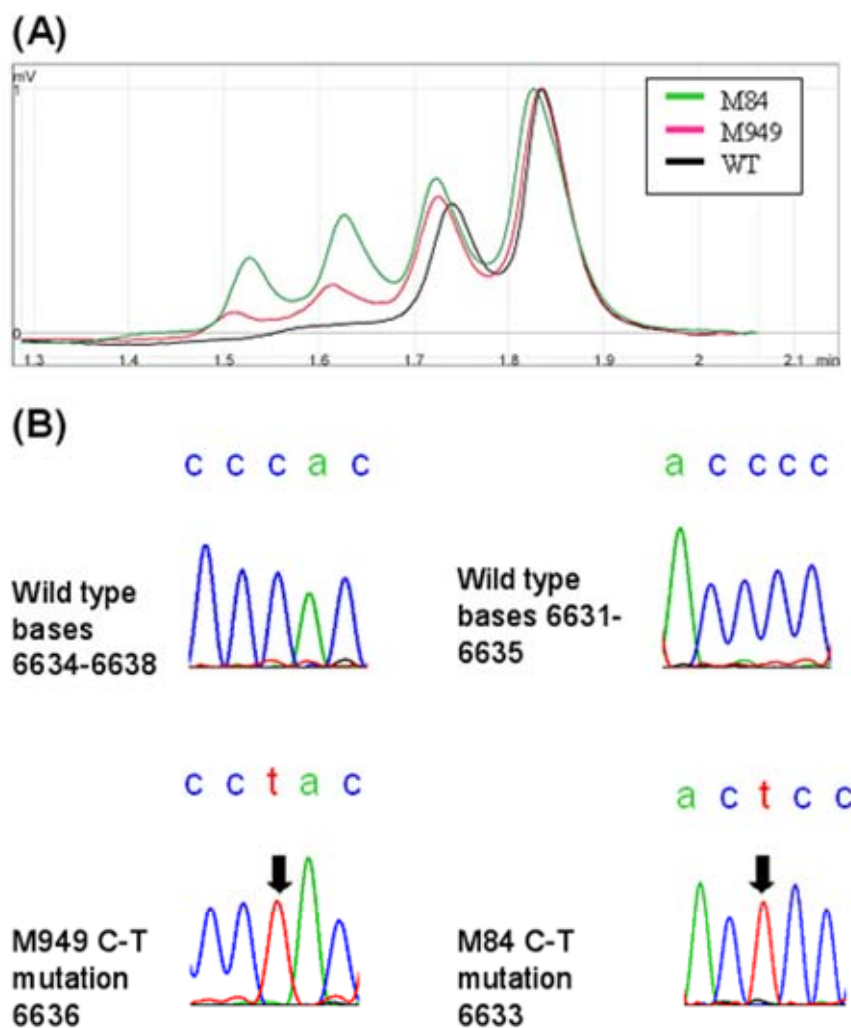


**Fig 6.6 Chromatogram traces from WAVE mutational analysis**

Representative traces (a) to (g) are shown for all coding exons. The second set of elution temperatures produced similar results identifying the positive control (OxMT) and the three abnormal samples (M949BC, M84BC & FMBC), marked by arrows, with no additional abnormal chromatograms. The legend lists the samples tested and their corresponding trace colour.

### 6.3.4.3 Sequence analysis of samples with abnormal products

To ascertain the cause of the abnormal chromatograms, samples M949BC, M84BC and FMBC along with 3 other randomly chosen control sequences with normal chromatograms (M997BC, M563BC, DTBC) and HEK293 genomic DNA underwent sequencing of all amplified fragments. Initially residual PCR products (35µl) were cleaned up using a Qiagen QIAquick PCR purification kit. After suspension in 30µl of elution buffer, 4µl of product was used in a BigDye terminator sequencing reaction v3.1 (Applied Biosciences) using bi-directional primers (as described in section 2.7.6). The sequencing reaction was run overnight, the products were then ethanol precipitated and then run on an ABI 3100 capillary array sequencer (Perkin Elmer, Beaconsfield UK). Sequences were compared to the published GATA-1 sequence with annotation for known single nucleotide polymorphisms found in the Ensembl Human-Gene SNPView database ([www.ensembl.org/Homosapiens/genesnpview](http://www.ensembl.org/Homosapiens/genesnpview)).



**Fig 6.7 WAVE and sequencing chromatograms for the exon 5 fragment of samples M949BC and M84BC compared to wild-type DNA.**

This analysis confirmed a mutation in M949BC and M84BC but in none of the other samples. M949BC had a single nucleotide substitution (C>T) in the intronic sequence 43 basepairs upstream of exon 5. M84BC had a single nucleotide substitution (C>T) 40 basepairs upstream of exon 5. The WAVE and sequencing chromatograms are shown in Fig 6.7(a) and Fig 6.7(b) respectively. These sequence variations do not correspond to known SNPs.

### **6.3.5 Verification of results using TOPO cloning of PCR products**

To confirm these results and ensure they were not due to mutations introduced by PCR amplification these two samples were independently re-analysed. PCR of exon 5 was performed using AmpliTaq gold and Pwo polymerase under standard conditions (section 2.10.1) followed by gel electrophoresis. Only one band was produced and this was excised, gel purified (Qiagen gel extraction kit) and eluted in TE buffer. Since use of a proof reading DNA polymerase with 3'-5' exonuclease activity results in blunt-ended PCR products it is necessary to add "A" overhangs before TOPO cloning. This was achieved by addition of dATP and Taq polymerase to the reaction mixture as detailed in section 2.12.1.1. The resultant product was cloned into a pCR4 sequencing vector (Invitrogen) as described in section 2.12.1.1. After transformation into DH5 $\alpha$  cells and growth overnight in LB medium in the presence of Ampicillin, cultures were lysed and DNA extracted using a Qiagen miniprep kit (section 2.12.2.3). Miniprep DNA (concentration 200ng/ $\mu$ l) was sent for commercial Sanger sequencing (Agowa.de). This sequencing confirmed the above mutations and did not reveal any additional changes.

## **6.4 Summary**

The results presented above show that GATA-1 sequences in public databases, obtained from patients with CML commonly show a higher rate of sequence variation than sequences obtained from other sources. In addition the analysis of the BC009797 clone described in chapter 3 revealed a mutation producing a change in the coding portion of exon 6 and altered functional activity as measured by an in vitro luciferase reporter assay. To assess whether GATA-1 mutations are a common feature in chronic myeloid leukaemia, mutational analysis of 24 patients with CML was performed. This analysis revealed a mutation rate of 2/24 but these mutations were confined to non-coding regions. The possible impact of non-coding sequence variations on gene expression are discussed in chapter 7. Overall, it seems unlikely that GATA-1 mutations play a major role in the

pathogenesis of CML blast crisis although this does not rule out an occasional role. In addition these experiments have not ruled out a role for aberrant regulation of GATA-1 expression in this disease, this has yet to be formally tested.

## 7 Conclusions and future directions

---

### 7.1 Summary of findings

This thesis set out to investigate the roles of GATA-1 isoforms in normal and malignant haematopoiesis and in particular to focus on the GATA-1s isoform and any unique properties it might display. The major findings from the previous chapters are summarised below:

1. The GATA-1s isoform is widely expressed in the same haematopoietic tissues as the GATA-1FL isoform and at the same stages of development. Unlike previously published reports no evidence was obtained for GATA-1s expression preceding GATA-1FL expression in the early developing embryo. There was a suggestion that GATA-1s may be preferentially expressed in one monocytic primary cell population but other monocytic cell lines and primary samples showed both GATA-1s and GATA-1FL expression. This observation needs further investigation.
2. Comprehensive examination of published databases and GATA-1 PCR products from various tissues revealed evidence of an intermediate splicing variant but no other physiological isoforms of GATA-1. However these investigations did reveal the existence of a mRNA with an altered C terminus, cloned from a patient with CML. Further investigation of this clone showed that it exhibited increased transactivation potential compared to human GATA-1FL in a luciferase reporter assay.
3. Attempts to produce GATA-1 isoform specific gene knockouts in embryonic stem cells met with technical difficulties, despite apparently successful targeting on PCR screening of colonies.
4. Ectopic over-expression of the GATA-1 isoforms in murine ES cells revealed that GATA-1FL promotes BFU-E and CFU-GM colony formation whilst GATA-1s over-expression significantly reduces haematopoietic colony output.
5. Ectopic expression in human meg-erythroid cell lines reveals differential roles for the two isoforms in driving erythroid and megakaryocytic commitment. GATA-1FL drives increased erythroid gene expression in K562 cells. On induction of megakaryocytic differentiation (with TPA) GATA-1FL produces increased DNA ploidy and suppression of

erythroid genes. GATA-1s acts as a hypomorphic allele in most situations, unable to activate or repress genes to the same extent as GATA-1FL. However GATA-1s does show upregulation of *ski* – a gene associated with bipotent meg-erythroid precursors which requires down-regulation prior to terminal erythroid differentiation. Although generally producing more repression of erythroid genes than megakaryocytic genes the inability of GATA-1s to support terminal megakaryocytic differentiation is suggested by failure to up-regulate Gp1b levels during TPA induced differentiation.

6. Comparison between trisomy 21 containing cells and disomic cells did not reveal any clear mechanistic reasons for the tight association between GATA-1s producing mutations and trisomy 21. However the presumed selective advantage of this association was strengthened by the observation that a trisomy 21 containing cell line (Meg-01) spontaneously selected for high level GATA-1s expression in culture, whilst a disomic cell line (K562) selected for high level GATA-1FL expression.

7. Further investigation of the possible association between GATA-1 mutations and malignancy did not reveal coding mutations in GATA-1 in blast cell genomic DNA isolated from a cohort of 21 patients with CML blast crisis. Two single base pair alterations in intronic sequences were discovered.

The potential significance of these findings and how they fit into the current body of knowledge on GATA-1 along with possible future lines of investigation are discussed in the following sections:

## **7.2 How do these findings extend current knowledge of GATA-1 function?**

These investigations aimed to shed light on two main areas: firstly the physiological role of the GATA-1s isoform and secondly the role it plays in TAM/AMKL.

### ***7.2.1 The physiological role of GATA-1s***

To date very little is known about the physiological function of GATA-1s. By inference from the phenotype of GATA-1s knock-in mice and DS-TAM/AMKL it appears that any physiological role is likely to occur in megakaryocytic or erythroid cells and may be restricted to early in haematopoietic ontogeny since both the mice and the human subjects

show defects in fetal haematopoiesis (Li, *et al* 2005, Vyas and Roberts 2006). However, this phenotype may be due to lack of the GATA-1FL isoform rather than the presence of GATA-1s in these cells. A number of hypotheses can be put forward as to the physiological function of GATA-1s; evidence from this thesis, in support of or against these hypotheses will be discussed:

**1. *GATA-1s may play a unique role in the determination of a certain cell lineage:*** Up until now it was unclear as to whether GATA-1s expression was restricted to certain lineages or developmental stages. If so, this pattern of expression might shed light on its physiological role. The findings presented in this thesis suggest that GATA-1s almost always accompanies GATA-1FL expression (with the possible exception of one monocytic cell preparation). This argues against a dominant role for GATA-1s in driving a particular developmental pathway. In addition, confirmation that mice do not produce this isoform by alternative splicing, but instead appear to ensure low level production by “leaky scanning”, make it highly unlikely that the GATA-1s isoform would play a unique role in the absence of GATA-1FL since it is difficult to envisage a mechanism whereby leaky scanning could be controlled sufficiently to prevent GATA-1FL production but allow GATA-1s production. Although the possibility of differential post-translational control in different cell types exists, western blot data suggest that again GATA-1s and GATA-1FL expression usually accompany one another at the protein level. Therefore, the body of evidence presented in this thesis argues against GATA-1FL and GATA-1s acting in a classical “alternative isoform” manner i.e. the switching between isoform usage in different cells or different developmental stages to produce different transcriptional outputs (as is seen with other transcription factors such as CEBP $\alpha$  (Ford, *et al* 1996)).

**2. *GATA-1s may compete with GATA-1FL for DNA or co-factor binding and promote cellular proliferation rather than differentiation:*** Existing microarray data from GATA-1s mice (Li, *et al* 2005) suggest that GATA-1s can activate the same set of genes as GATA-1FL but that a subset of these genes are inadequately repressed or up-regulated in the presence of GATA-1s. This is particularly true of genes associated with cell proliferation (such as c-myb, c-myc, cyclin D2) or alternative lineage choice (PU.1, Ikaros, GATA-2). This suggests that GATA-1s is capable of mimicking some of the functions of GATA-1FL but lacks the ability to promote lineage switching or terminal differentiation at physiological levels. Its particularly important role in promoting cellular proliferation at the expense of differentiation is supported by GATA-1 rescue experiments discussed in section 1.2.5.3 where GATA-1 N-terminal mutants could rescue the differentiation block but not the abnormal proliferation of GATA-1 knockout cells (Kuhl, *et al* 2005, Muntean

and Crispino 2005). These results tie in with cell line studies showing that GATA-1 effects on proliferation and differentiation appear to occur independently. For example, restoration of GATA-1 expression in the GATA-1 null cell line G1E allows terminal differentiation (Weiss, *et al* 1997). This has been shown to be associated with G1 cell cycle arrest associated with down-regulation of Cyclin D2 levels and due, at least in part, to binding to and repression of the c-myc promoter by GATA-1 (Rylski, *et al* 2003). Over expression of c-myc in these cells prevented the cell cycle arrest but had minimal effects on restoring maturation. Other studies, using a tetracycline inducible conditional GATA-1 expression plasmid in knockout ES cells, showed that only short windows of GATA-1 expression early during in vitro differentiation were needed for rescue of proliferation defects whilst more sustained and later expression was needed for effective differentiation (Zheng, *et al* 2006).

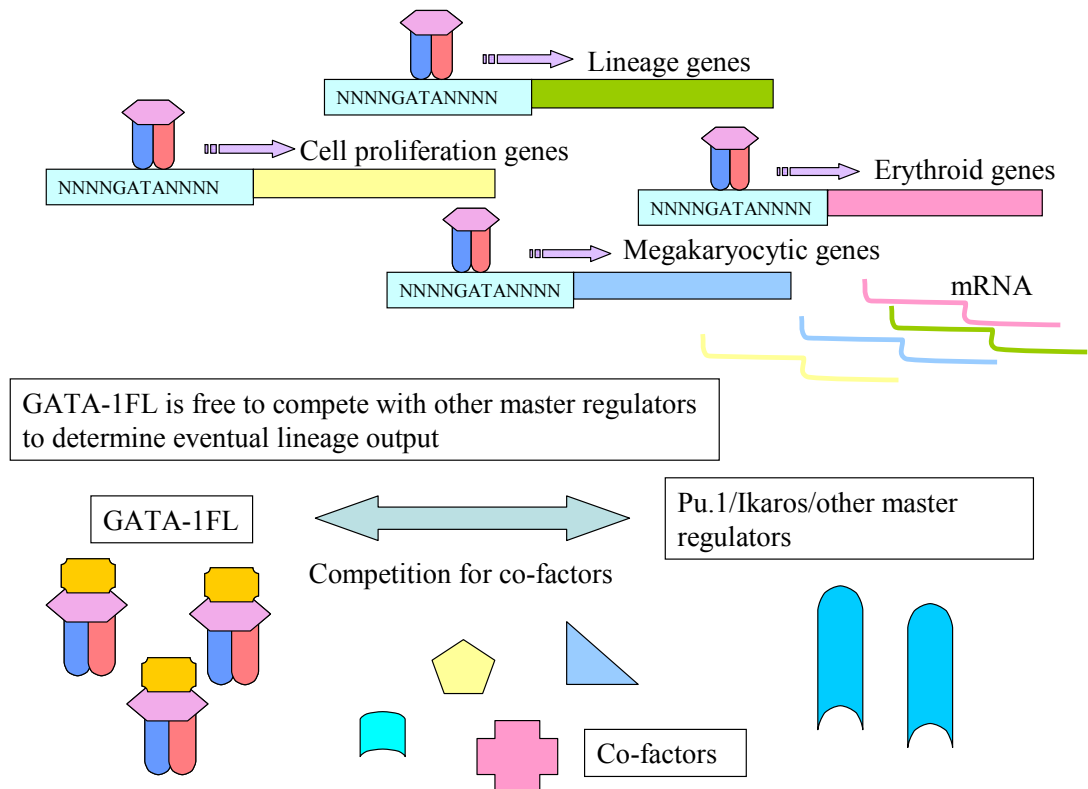
Therefore it could be postulated that GATA-1s is expressed at a basal level and prevents repression of genes responsible for cell cycling. The fact that it also fails to repress alternative lineage specific genes suggests that it may act in fairly primitive HSCs and promote maintenance of these multi-potential cells in a primed but proliferative state ready for terminal maturation decisions (illustrated in fig 7.1a). GATA-1FL (and other master regulators such as PU.1) would then be free to interact with all the necessary co-factors and transcriptional partners in the cell, the balance of which would determine eventual cell fate. It would be envisaged that these transcription factor complexes would then compete with, and replace, GATA-1s binding at the various loci to produce terminal differentiation (as illustrated in Fig 7.1b). A similar model, whereby haematopoietic precursors exist in a primed state expressing genes from multiple lineages before commitment to a single fate, has been proposed (Hu, *et al* 1997) and is supported by demonstration of mRNA level expression of multiple, supposedly lineage specific, genes within a single haematopoietic progenitor cell.

It is interesting that another key haematopoietic transcription factor AML-1 (Runx-1) also possesses a number of isoforms one of which (AML-1a) is very similar structurally to GATA-1s in that it contains the AML-1 DNA binding domain but lacks the transcriptional regulatory domain and has a reduced transactivation potential. Recent studies over-expressing this isoform (AML-1a) in haematopoietic stem cells have shown that it promotes proliferation of progenitors, greatly increasing engraftment in an LTR-HSC assay and increasing primitive and committed progenitor activity. In contrast the full-length isoform (AML-1b) promoted terminal differentiation (Tsuzuki, *et al* 2007). It seems entirely plausible that GATA-1, another “master regulator” of haematopoietic cell fate

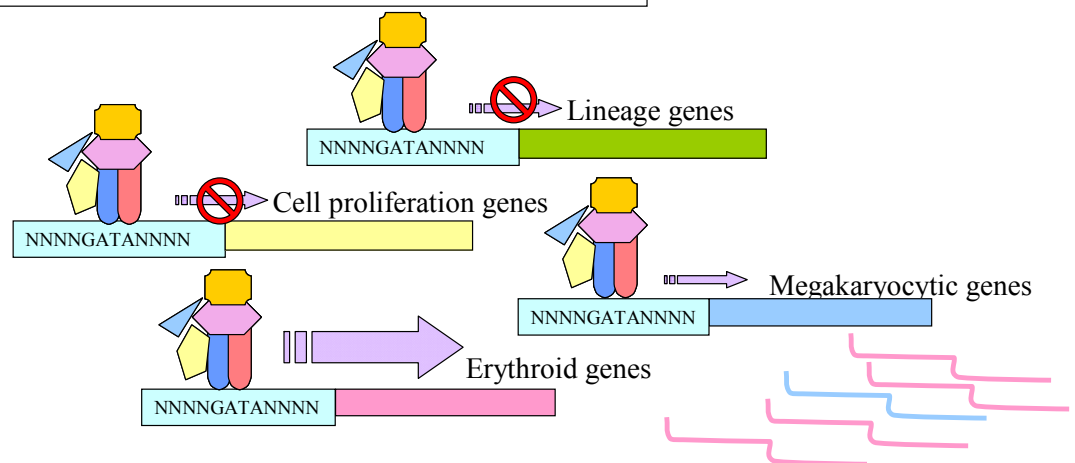


might employ similar mechanisms, and this raises the interesting possibility that this could be a generic method for coupling the ability to maintain a suitable primed progenitor compartment with the ability to induce terminal differentiation – of note other master regulators such as PU.1 have multiple isoforms, the individual roles of which are unclear.

- (a) GATA-1s binds upstream regulatory elements in uncommitted progenitors activating basal transcription to allow lineage priming and expansion of this compartment

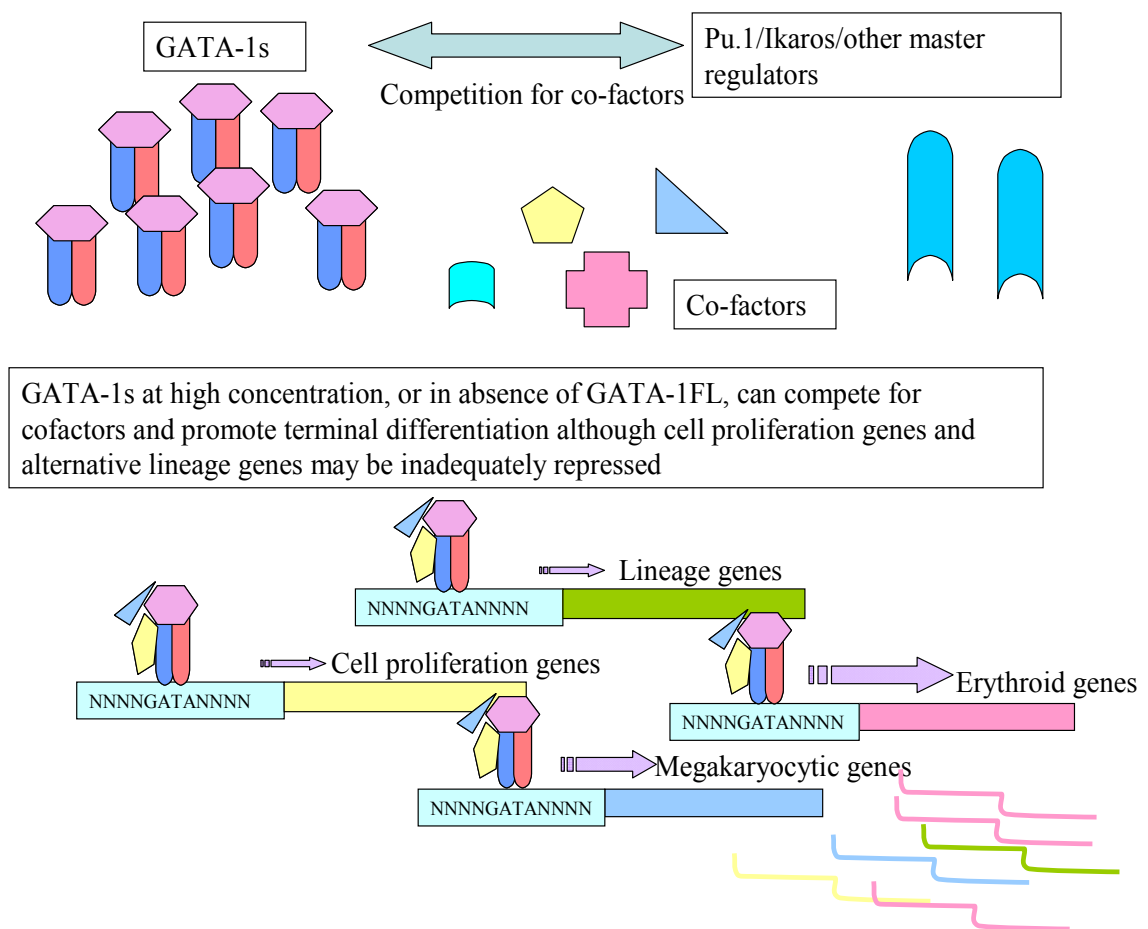


- (b) GATA-1FL containing transcription factor complexes displace GATA-1s to permit terminal differentiation



**Fig 7.1 A potential role for GATA-1s in haematopoietic progenitor cells**

This hypothesis must be able to explain the normal terminal maturation seen if the GATA-1s isoform is over expressed (Muntean and Crispino 2005, Shimizu, *et al* 2001). Since a large number of GATA-1 co-factors and transcription factors bind to the Zinc fingers (FOG-1, AML-1, ski etc) and not the N-terminal domain, GATA-1s expressed at supra-physiological levels might be able to compensate for absence of GATA-1FL and would tip the balance in favour of GATA-1 terminal differentiation (as opposed to alternative lineage outputs). There may however, still be low level detection of alternative lineage genes and failure to repress pro-proliferative genes resulting in expansion of precursors. This is illustrated in Fig 7.2.



**Fig 7.2 High levels of GATA-1s may be able to compensate for loss of GATA-1FL**

A number of pieces of evidence from this thesis support aspects of this hypothesis i.e. the gene expression studies show that GATA-1s generally acts as a hypomorphic allele compared to GATA-1FL during terminal differentiation. The higher levels of expression of cell cycle and alternative lineage genes with transgenic GATA-1s expression, compared to GATA-1FL expression, are confirmed. In general GATA-1s expression is associated with reduced terminal maturation (e.g. lower DNA ploidy). However, when GATA-1s is expressed at higher levels (as in the Meg-01 GATA-1s transfectants) it is associated with

increased ability to activate or repress target genes almost mimicking the GATA-1FL phenotype. Further testing of this hypothesis is required as discussed in section 7.3 below.

### **3. *GATA-1s may play a particular role in primitive haematopoiesis***

Early on in haematopoietic ontogeny rapid expansion of primitive haematopoietic precursors with a skew towards erythroid and megakaryocytic development occurs. In this context expression of the GATA-1s isoform has been shown to fully compensate for lack of GATA-1FL even when expressed at physiological levels (Shimizu, *et al* 2001). Since this subset of cells has much more limited developmental potential (i.e. primitive HSCs are unable to reconstitute the lymphoid compartment and have restricted myeloid potential) the need to interact with multiple co-factors may be reduced and DNA binding by GATA-1FL or GATA-1s may be sufficient to induce erythroid and megakaryocytic differentiation. The fact that GATA-1s promotes proliferation may be advantageous in this context by allowing rapid expansion of the progenitor pool. This role in primitive haematopoiesis is not mutually exclusive with the role in priming progenitors discussed above. It explains the particular propensity for GATA-1s mutations to affect fetal liver haematopoiesis but still permit terminal differentiation (in the absence of trisomy 21) (Li, *et al* 2005).

This role in primitive haematopoiesis is supported by results presented in section 3.2.5 showing GATA-1s expression at the protein level in murine embryonic tissues including the E9 yolk-sac. Studies presented here do not confirm an isolated report in the literature of GATA-1s production preceding GATA-1FL early in murine development (Calligaris, *et al* 1995). As already discussed above, the isolated production of GATA-1s in mice would have to entail a complicated mechanism of “leaky scanning” favouring translation of the second ATG, or (probably more likely) differential post-translational regulation for the two isoforms (i.e. by modifications targeting the N-terminus). In addition, the balance of gene expression studies presented here, suggest that GATA-1s tends to favour megakaryocytic gene expression over erythroid. This taken together with the megakaryocytic phenotype in TAM/AMKL and GATA-1s knock-in mice suggests that GATA-1s may be important in the recently described primitive megakaryopoiesis detected in early murine embryos (Tober, *et al* 2007). This is also supported by the relatively “primitive” nature of the platelets (i.e. discoid, large and reticulated) seen in the human kindred with a GATA-1s producing mutation (Hollanda, *et al* 2006). Again this hypothesis remains to be tested.

### **7.2.2 The role of GATA-1s in TAM/AMKL**

Obviously understanding the physiological role of GATA-1s is of fundamental biological importance and may help shed light on its “non-physiological” role in malignancy. In addition, understanding the role of GATA-1s in haematological malignancy is important for a number of reasons:

1. GATA-1s associated malignancies develop in a step-wise manner from an abnormal meg-erythroid compartment in trisomy 21; to emergence of TAM with GATA-1s mutations; followed by either resolution or acquisition of additional mutations to produce AMKL. Dissection of each stage of this process provides a rare opportunity to unravel the multiple processes involved in acquiring a haematological malignancy using an *in vivo* biological model with distinct pre-malignant stages.
2. The fact that TAM often spontaneously resolves offers hope that understanding the mechanisms of GATA-1s cellular transformation may allow insight into how to reverse a malignant phenotype by promoting clonal extinction. Although this may be predominantly of use in paediatric malignancies, affecting lineages that occur during certain developmental windows, the observation (confirmed in this thesis) of reversion to fetal type haematopoiesis in cell lines derived from some adult malignancies (such as K562), suggests that it may have wider relevance.
3. Knowledge of the effects of GATA-1s expression, its interacting partners in the cell and the mechanism by which it promotes a malignant phenotype may be of direct benefit to children with DS-TAM/AMKL, either by suggesting novel therapeutic strategies or allowing better prediction of which patients can safely be observed and which need early therapeutic intervention.

As discussed in chapter 1 it is thought that GATA-1s contributes to the TAM/AMKL phenotype by promoting expansion/immortalisation of meg-erythroid progenitor cells in the fetal liver (a compartment already expanded in the presence of trisomy 21). How it does this is unclear, although its inability to suppress *c-myb*, *c-myc*, cyclins and cyclin dependent kinases has been suggested to be of importance (Li, *et al* 2005). The addition of trisomy 21 appears to be essential for the phenotype (GATA-1s associated malignancy is only seen in the presence of trisomy 21). *Erg3*, *Ets2* and *Runx-1* are all chromosome 21 genes which are candidates for driving expansion of the meg-erythroid compartment and another chromosome 21 gene- *Bach-1* (over-expressed in DS-AMKL and TAM in a

number of independent studies) may be important for blocking terminal differentiation (Rainis, *et al* 2005).

Results presented in chapter 5 confirm higher levels of c-myb and cyclin D2 in K562 GATA-1s transgenics and higher levels of c-myb in Meg-01 GATA1s transgenics, compared to GATA-1FL transgenics. This confirms a likely role for these genes but their mechanism of action is unclear. Unlike observations in the GATA-1s transgenic mice (Li, *et al* 2005), these human GATA-1s transgenic cell lines did not exhibit increased proliferation compared to GATA-1FL. c-myb in particular has a very complex role in haematopoiesis promoting both stem cell self-renewal and lineage decisions (Sakamoto, *et al* 2006). Therefore expression is likely to produce very different consequences depending on its exact cellular context. Since Meg-01 and K562 are both malignant human cell lines, it is not surprising that their phenotype might differ from murine megakaryocytic progenitors. Another interesting gene to emerge from this data is ski. As discussed in chapter 5 expression of ski appears to mark a bipotent meg-erythroid progenitor. Data presented here suggest that at baseline and during TPA and Haemin induced differentiation ski gene expression is significantly lower in the GATA-1FL transgenics compared to the GATA-1s and empty vector transgenics in K562 cells with a similar (but non-significant) trend in Meg-01 cells. Little is known about the mechanisms of action of ski in normal haematopoiesis and the literature suggests both that its presence is important for megakaryocytic differentiation (Namciu, *et al* 1994), and that its down regulation is needed for terminal differentiation (Pearson-White, *et al* 1995). Clearly further work is needed and these very preliminary data suggest that examination of ski expression levels during in vitro haematopoietic differentiation in other systems, and in TAM/AMKL cells, would be interesting.

The data from Meg-01 stable transfectants also supports the hypothesis that GATA-1s may display a unique selective advantage in the presence of trisomy 21. The consistent selection for high level GATA-1s expression in this cell line (and low level expression in K562) is intriguing and does not appear to reflect a simple proliferative advantage. Unfortunately the multiple differences between the K562 and Meg-01 cell lines in terms of additional genetic abnormalities, stage of haematopoietic commitment etc make it difficult to tease out the reasons behind this selective pressure using this experimental system. Alternative approaches are discussed in section 7.3 below.

Finally, the expression of two candidate chromosome 21 genes – Erg3 and Bach-1- was assessed in this thesis. Erg3 was detectable in Meg-01 cells but not K562 consistent with

its over-expression in trisomy 21 although the different levels of haematopoietic commitment may also be responsible. Erg3 was significantly up-regulated on TPA and Haemin treatment of both cell lines (although it remained persistently higher in Meg-01 cells) but showed no differential expression patterns between the 3 transgenes. This is not surprising as Erg3 is not postulated to be a GATA-1 target gene and presumably acts upstream of GATA-1 to cause a pre-existing expansion of the meg-erythroid compartment. Therefore, from this data Erg3 remains a plausible candidate whose over-expression in trisomy 21 is a possible driver of the expansion of the meg-erythroid pool. Relevant to this a recent paper described the effects of ectopic over-expression of Erg3 (and Ets2) in GATA-1s transgenic ES cells, wild-type ES cells or the megakaryocyte restricted GATA-1 knockdown ES cells ( $\Delta$ neo $\Delta$ HS – discussed in chapter 1 section 1.2.5.1) (Stankiewicz and Crispino 2009). Erg3 was able to promote immortalisation of GATA-1s transgenic ES cells but did not produce a fully malignant phenotype, as assessed by ability to produce disease when transplanted into mice. Therefore, unsurprisingly, it is unlikely that Erg3 is the sole partner for GATA-1s in producing the TAM phenotype.

Gene expression data from Meg-01 and K562 cells, discussed in chapter 5, do not support a role for over-expression of Bach-1 in the context of trisomy 21. Meg-01 cells and K562 cells showed almost identical baseline expression of Bach-1 and dynamic modulation during TPA and Haemin treatment showed increased up-regulation of Bach-1 gene expression in K562, with appropriate down-regulation in Meg-01. In addition, the only statistically significant difference between the transgenes showed increased repression of Bach-1 by GATA-1s during Haemin differentiation in Meg-01. This suggests that, even in the context of GATA-1s expression, Bach-1 is not over-expressed. Of course, there are a number of explanations for these data. The use of comparative fold-change rather than absolute quantitation of cDNA levels means it is equally possible that K562 cells show up-regulation of Bach-1 (by mechanisms other than trisomy 21), rather than Meg-01 showing normal levels of expression, comparison with an equivalent normal cell population would be useful here. Alternatively, Bach-1 over-expression may be restricted to a particular stage during haematopoietic ontogeny, due to the need for co-factors or permissive epigenetic factors, and may be normal in haematopoietic cells derived from adults even in the presence of trisomy 21 (this hypothesis would provide an explanation for the developmentally restricted phenotype of GATA-1s knockout mice, but remains to be tested).

***Overall, results from this thesis support a role for GATA-1s in promoting cellular proliferation and survival, inadequately repressing genes associated with alternative***

*lineage choice and generally favouring megakaryocytic gene expression over erythroid. In addition GATA-1s expression seems to produce a selective advantage in cells carrying trisomy 21. All of these features are consistent with outgrowth of a primitive meg-erythroid precursor in fetal liver cells carrying a GATA-1s producing mutation, with exacerbation of this phenotype in the presence of trisomy 21. In addition they confirm that GATA-1s can support terminal differentiation if over-expressed at high enough levels. Finally a role for Erg3 in promoting meg-erythroid differentiation in trisomy 21 cells is supported but the role of Bach-1 in suppressing terminal differentiation could not be confirmed.*

### **7.2.3 The role of GATA-1 in other malignancies**

Although GATA-1 mutations were thought to be uniquely associated with DS-TAM and AMKL (Wechsler, *et al* 2002), the association between GATA-1 mutations and CML has never previously been investigated. As discussed in chapter 1 and Chapter 6, a number of observations suggested that perturbations in GATA-1 function may be important in the pathogenesis of CML. This thesis provides data showing that at least one CML patient expressed an abnormal GATA-1mRNA transcript characterised by shortening and frameshift in the final exon involving the C-terminal tail. Interestingly this mutant cDNA showed increased transactivation in a luciferase reporter assay compared to human GATA-1FL and GATA-1s. The distal C-terminal tail of GATA-1 has not been extensively characterised but interestingly appears to bind AML-1 (Elagib, *et al* 2003, Waltzer, *et al* 2003) and the conserved QTRNRK motif, found just distal to the C-terminal zinc finger, has been demonstrated to be important for determining sequence specificity of GATA-1 binding (Ghirlando and Trainor 2003). In addition a recent structural analysis of a GATA-1 homologue, MED-1, found in *Caenorhabditis elegans*, has shown that this GATA-type zinc finger domain possesses unique binding specificity by virtue of basic residues in the tail region of the molecule distal to the C-terminal zinc-finger (Lowry, *et al* 2008). These residues allow the tail to take up an  $\alpha$ -helical structure and interact with the DNA major binding groove producing an altered and extended DNA consensus sequence. It is also known that the distal C-terminal tail contains residues that can be post-translationally modified by acetylation or phosphorylation (Boyes, *et al* 1998, Hung, *et al* 1999). Experiments using site-directed mutagenesis of terminal lysine residues in chicken GATA-1 showed that alteration in acetylation patterns can alter DNA binding and transactivation potential (Boyes, *et al* 1998). The C-terminal mutant described in this thesis displays altered transactivation potential despite an intact N-terminus and N- and C-terminal zinc

fingers suggesting that the distal tail is responsible for this phenomenon. Interestingly the mutant frameshift occurs proximal to the important QTRNRK motif (Ghirlando and Trainor 2003). One exciting possibility is that the modifications to the C-terminal tail of GATA-1 (be it mutations, binding of co-factors or physiological acetylation or phosphorylation) could alter its conformation and thereby alter DNA-binding specificity. This could provide an additional mechanism whereby GATA-1 could switch between active and repressive states or produce a different pattern of gene expression in different lineages. This hypothesis needs testing.

Further investigations of GATA-1 mutations in CML did not reveal any additional mutations in coding regions, two single base-pair mutations were seen approximately 40bp upstream of the 5<sup>th</sup> exon. Although these may represent non-functional mutations or previously unrecognised SNPs it is also possible that these mutations are of functional significance. The role of intronic sequences in the regulation of gene expression is increasingly being appreciated both as potential microRNA precursors and in control of splicing. One possibility is that these mutations affect the branch-point consensus sequence required for lariat formation during intronic splicing. The human branch-point sequence is quite variable and generally located around 35 basepairs upstream of the 3' end of the intron, (Gao, *et al* 2008), this corresponds to the mutated region in these studies. Further analysis of the presence of aberrantly spliced transcripts in these patients would be interesting, but unfortunately is hampered by unavailability of RNA in these cases.

Finally, the lack of coding mutations in GATA-1 in this cohort of CML patients does not rule out a functional role for GATA-1 in this condition. Since correct modulation of GATA-1 expression levels is critical for terminal differentiation it is possible that aberrant GATA-1 expression is just as important as mutation of the protein. There is some evidence of an interplay between bcr-abl and GATA-1 for example an investigation of the mechanism of upregulation of inducible heat shock protein 70 (Hsp70) in CML showed that the major p210 bcr-abl responsive element upstream of the Hsp70 promoter mapped to a GATA-1 binding motif (Ray, *et al* 2004).

***In conclusion, these findings suggest that GATA-1 mutations are not frequent occurrences in the evolution of blast crisis in CML. However, this does not rule out an important role for deregulated GATA-1 activity in the molecular pathogenesis of CML. In addition, investigation of a CML derived abnormal mRNA transcript suggests that the GATA-1 C-terminal tail may be an important, previously under-appreciated functional region of the molecule.***



## 7.3 Future investigations

The results presented here throw up a number of potential avenues for further investigation. In particular:

### **1. Is the GATA-1s expression in monocytes physiologically relevant or an artefact?**

The monocytic primary human cell sample described in section 3.2.1 remains the only tissue in which GATA-1s expression was seen in the absence of GATA-1FL. However, testing of another primary sample and a number of monocytic cell lines failed to confirm these findings. Despite this some pieces of circumstantial evidence suggest a possible role for GATA-1s in monocytic differentiation. In particular levels of D6 have been shown to be upregulated during monocytic differentiation into dendritic cells in a GATA-1 dependent manner (McKimmie, *et al* 2008). Interestingly, D6 was significantly upregulated by the GATA-1s isoform compared to GATA-1FL in the K562 transgenic experiments reported in chapter 5 (Fig 5.15b). Overall these observations suggest that additional study of isoform expression in monocytes and dendritic cells might be informative.

### **2. Testing the hypothesis that GATA-1s is associated with priming haematopoietic precursors.**

This interesting possibility could be further tested by transfecting primitive haematopoietic cells with GATA-1s and GATA-1FL expression plasmids and testing their behaviour in long-term reconstitution assays and colony assays. The real test of whether GATA-1s has a unique role in haematopoietic development (i.e. rather than just acting as a dispensable hypomorphic allele of GATA-1FL) would be provided by the generation and testing of a GATA-1s knock-out (as opposed to the GATA-1 “knock-in” already described in the literature (Li, *et al* 2005)). This was attempted in this thesis, without success, although the feasibility of mutating the second start codon without detriment to GATA-1FL function was proven. It would still be worthwhile to pursue this aim. Huge advances in gene targeting technology have been achieved in the last 5 years (since the inception of this PhD thesis) in particular the availability of recombineering technology allows rapid cloning of targeting constructs from bacterial artificial chromosomes (BACs) using bacterial recombinases (Wu, *et al* 2008). This reduces the time taken to make a targeting construct from a number of months to a couple of weeks and therefore facilitates multiple attempts at producing an effective construct. It is well known that some genes are much more difficult to target than others, due to factors such as sections of repetitive DNA, significant secondary and tertiary structure or other unknown factors. Successful targeting

can often be achieved in these cases by altering the locations of the homology arms (Wu, *et al* 2008). This is often impractical using conventional cloning due to the time involved, but recombineering makes multiple attempts at targeting vector construction (or parallel attempts with a number of different vectors) feasible. In addition one of the major factors reducing targeting efficiency is lack of complete sequence homology. The use of BACs overcomes potential errors introduced by PCR amplification of homology arms and prevents exhaustive sequencing of PCR products and attempts to decipher whether sequence variations represent SNPs, genuine sequence variation or errors introduced by the DNA polymerase.

Other approaches to assessing whether the GATA-1s isoform has a unique role in haematopoiesis include assessing whether it binds to a different subset of genes than GATA-1FL, at defined stages of ontogeny, or within defined cell lines, at both the level of DNA binding at gene regulatory elements and binding to different protein co-factors. Assessment of DNA binding at promoters, enhancers etc could be achieved using chromatin immunoprecipitation (ChIP) studies. Whereas binding to co-factors could be assessed by co-immunoprecipitation studies followed by mass spectrometry to identify co-purifying protein partners. These investigations are made more difficult by the fact that GATA-1s doesn't contain any unique domains and therefore antibodies to GATA-1s will also bind to GATA-1FL. This could be overcome in one of two ways. Firstly, a kind of "subtractive" ChIP or co-immunoprecipitation could be performed exploiting the fact that the N6 antibody will only bind to GATA-1FL whilst the M20 antibody will bind to both isoforms, therefore results using these two antibodies could be compared and differences ascribed to the GATA-1s isoform. Secondly, a better approach might be to tag the GATA-1s isoform, allowing precipitation using antibodies to the tag rather than GATA-1s. This could be achieved either using a single tag such as FLAG or biotin or using a tandem affinity purification (TAP) approach which improves specificity and purity of protein complexes by two successive selection procedures (Knuesel, *et al* 2003).

**3. Testing the role of GATA-1s in primitive megakaryopoiesis.** The existence of primitive megakaryopoiesis has only recently been described (Tober, *et al* 2007). This study used immuno-histochemical staining of embryonic tissues and culture of yolk-sac and embryo derived haematopoietic progenitor cells. In order to test the role of GATA-1s it would be very interesting to repeat these early progenitor assays in GATA-1s knock-in mice. Again, the real test of the role of GATA-1s would be to look at primitive megakaryopoiesis in a GATA-1s isoform specific knockout as discussed above. Therefore further attempts to generate these cells would also be worthwhile in this context. Although

these experiments would require generation of transgenic mice to look at yolk-sac derived progenitors, there are also in vitro assays of megakaryopoiesis using murine ES cells grown on OP9 stromal cell layers (Eto, *et al* 2002). These assays are reported to show two waves of megakaryopoiesis with small immature megakaryocytes appearing on day 8 of differentiation and larger cells exhibiting hyperdiploidy appearing on day 12 (Fujimoto, *et al* 2003). This is highly suggestive of waves of primitive and definitive megakaryopoiesis, but this is yet to be proven. The use of GATA-1s and GATA-1FL knock-out ES cells in these assays would have enormous potential to address the hypothesis that GATA-1s plays a crucial role in primitive megakaryopoiesis.

**4. Further investigation of the role of GATA-1s in TAM/AMKL.** This thesis has suggested that it would be interesting to look at the expression of ski in TAM/AMKL samples and that studies to address the role of ski in megakaryocytic development are needed. In addition, an improved model for assessing the role of trisomy 21 is needed, Meg-01 cells are interesting, but interpretation is limited by their additional genetic changes and derivation from adult CML. The recent development of a transgenic mouse carrying an additional copy of human chromosome 21 (O'Doherty, *et al* 2005) holds much promise for dissecting out the role of chromosome 21 genes in the pathogenesis of this condition. The use of ES cells from these mice for in vitro haematopoietic differentiation assays would be extremely interesting, with or without additional manipulation of GATA-1s expression. Ultimately, it might be possible to cross breed these mice with GATA-1s transgenics to provide a better animal model of TAM/AMKL, although this might be hampered by lack of stability of the trisomic genotype.

**5. Further investigation of the C-terminal tail of GATA-1.** Finally, one of the most biologically interesting findings from this thesis is the identification of the distal C-terminal tail as a potentially important GATA-1 regulatory region. This could be further investigated in a number of ways. The importance of the tail could be assessed by creating C-terminal deletion mutants and assessing their function in various reporter assays or binding assays. The existing mutant could also be further investigated – the increased transactivation potential needs confirming using a different system. In addition, it would be interesting to look at the DNA binding profile of this mutant (by ChIP) and whether it is capable of interacting with AML-1 (by co-immunoprecipitation). It would also be extremely interesting to assess which other proteins bind in this region of the molecule. In terms of modulation of GATA-1 activity by post-translational modifications, it is already known that prevention of lysine acetylation in the C-terminal tail, as well as other regions, leads to reduction in GATA-1 activity in vitro (Boyes, *et al* 1998)) although the

physiological relevance of this remains unclear. Similar mutagenesis studies to prevent serine phosphorylation (Crossley and Orkin 1994) were unable to show an impact on in vitro erythroid differentiation in MEL cells. This does not rule out an important role for this type of modification in certain lineages or during ontogeny. This needs further investigation. If C-terminal modifications were identified that may influence DNA binding then analysis of the impact of these modifications on the structure of GATA-1 could be assessed either using predictive software programmes, nuclear magnetic resonance spectroscopy or by attempts to crystallise the protein and assess by X-ray crystallography.

*In conclusion, this thesis has produced a number of testable hypotheses concerning a wide range of potentially novel functions for the different domains of GATA-1 including a possible role for GATA-1s in haematopoietic progenitors and a role for the C-terminal tail in modulation of GATA-1 function. Rapid advances in molecular biological techniques should allow creation of improved systems in which to test these hypotheses and uncover novel roles for this key haematopoietic transcription factor, so vital in both physiological and malignant haematopoiesis.*

## 8 References

---

- Adams, G.B., Chabner, K.T., Alley, I.R., Olson, D.P., Szczepiorkowski, Z.M., Poznansky, M.C., Kos, C.H., Pollak, M.R., Brown, E.M. & Scadden, D.T. (2006) Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature*, **439**, 599-603.
- Adolfsson, J., Mansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C.T., Bryder, D., Yang, L., Borge, O.J., Thoren, L.A., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M. & Jacobsen, S.E. (2005) Identification of Flt3<sup>+</sup> lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*, **121**, 295-306.
- Ahmed, M., Sternberg, A., Hall, G., Thomas, A., Smith, O., O'Marcaigh, A., Wynn, R., Stevens, R., Addison, M., King, D., Stewart, B., Gibson, B., Roberts, I. & Vyas, P. (2004) Natural history of GATA1 mutations in Down syndrome. *Blood*, **103**, 2480-2489.
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I.L. (2000) A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, **404**, 193-197.
- Al-Kasim, F., Doyle, J.J., Massey, G.V., Weinstein, H.J. & Zipursky, A. (2002) Incidence and treatment of potentially lethal diseases in transient leukemia of Down syndrome: Pediatric Oncology Group Study. *Journal of Pediatric Hematology Oncology*, **24**, 9-13.
- Alberts, B. (2008) *Molecular biology of the cell. Reference edition*. Garland Science, New York, N.Y; Abingdon.
- Alitalo, R. (1990) Induced differentiation of K562 leukemia cells: a model for studies of gene expression in early megakaryoblasts. *Leukemia Research*, **14**, 501-514.
- Alter, B.P. (2002) Bone marrow failure syndromes in children. *Pediatric Clinics of North America*, **49**, 973-988.
- Andersson, L.C., Jokinen, M. & Gahmberg, C.G. (1979) Induction of erythroid differentiation in the human leukaemia cell line K562. *Nature*, **278**, 364-365.
- Anguita, E., Hughes, J., Heyworth, C., Blobel, G.A., Wood, W.G. & Higgs, D.R. (2004) Globin gene activation during haemopoiesis is driven by protein complexes nucleated by GATA-1 and GATA-2. *European Molecular Biology Organisation Journal*, **23**, 2841-2852.
- Arinobu, Y., Iwasaki, H., Gurish, M.F., Mizuno, S., Shigematsu, H., Ozawa, H., Tenen, D.G., Austen, K.F. & Akashi, K. (2005) Developmental checkpoints of the basophil/mast cell lineages in adult murine hematopoiesis. *Proceedings of the National Academy of Sciences U S A*, **102**, 18105-18110.
- Arron, J.R., Winslow, M.M., Polleri, A., Chang, C.P., Wu, H., Gao, X., Neilson, J.R., Chen, L., Heit, J.J., Kim, S.K., Yamasaki, N., Miyakawa, T., Francke, U., Graef, I.A. &

- Crabtree, G.R. (2006) NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. *Nature*, **441**, 595-600.
- Ashley, D.J. (1969) The two "hit" and multiple "hit" theories of carcinogenesis. *British Journal of Cancer*, **23**, 313-328.
- Awatramani, R., Soriano, P., Mai, J.J. & Dymecki, S. (2001) An Flp indicator mouse expressing alkaline phosphatase from the ROSA26 locus. *Nature Genetics*, **29**, 257-259.
- Baird, J.W., Ryan, K.M., Hayes, I., Hampson, L., Heyworth, C.M., Clark, A., Wootton, M., Ansell, J.D., Menzel, U., Hole, N. & Graham, G.J. (2001) Differentiating Embryonal Stem Cells Are a Rich Source of Haemopoietic Gene Products and Suggest Erythroid Preconditioning of Primitive Haemopoietic Stem Cells. *Journal of Biological Chemistry*, **276**, 9189-9198.
- Baron, M.H. & Maniatis, T. (1986) Rapid reprogramming of globin gene expression in transient heterokaryons. *Cell*, **46**, 591-602.
- Bedi, R., Du, J., Sharma, A.K., Gomes, I. & Ackerman, S.J. (2008) Human C/EBP $\epsilon$  activator and repressor isoforms differentially reprogram myeloid lineage commitment and differentiation. *Blood*, **2**, 317-327.
- Bertrand, J.Y., Desanti, G.E., Lo-Man, R., Leclerc, C., Cumano, A. & Golub, R. (2006) Fetal spleen stroma drives macrophage commitment. *Development*, **133**, 3619.
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes and Development*, **16**, 6-21.
- Blau, H.M. & Baltimore, D. (1991) Differentiation requires continuous regulation. *Journal of Cell Biology*, **112**, 781-783.
- Blobel, G.A., Simon, M.C. & Orkin, S.H. (1995) Rescue of GATA-1-deficient embryonic stem cells by heterologous GATA-binding proteins. *Molecular and Cellular Biology*, **15**, 626-633.
- Bonnet, D. & Dick, J.E. (1997) Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Medicine*, **3**, 730-737.
- Bottardi, S., Ghiam, A.F., Bergeron, F. & Milot, E. (2007) Lineage-specific transcription factors in multipotent hematopoietic progenitors: a little bit goes a long way. *Cell Cycle*, **6**, 1035-1039.
- Bottardi, S., Ross, J., Bourgoin, V., Fotouhi-Ardakani, N., Affar, E.B., Trudel, M. & Milot, E. (2009) Ikaros and GATA-1 combinatorial effect is required for silencing of human  $\gamma$ -globin genes. *Molecular and Cellular Biology*, **29**, 1526-37.
- Bouilloux, F., Juban, G., Cohet, N., Buet, D., Guyot, B., Vainchenker, W., Louache, F. & Morle, F. (2008) EKLF restricts megakaryocytic differentiation at the benefit of erythrocytic differentiation. *Blood*, **112**, 576-584.

- Bourquin, J.P., Subramanian, A., Langebrake, C., Reinhardt, D., Bernard, O., Ballerini, P., Baruchel, A., Cave, H., Dastugue, N., Hasle, H., Kaspers, G.L., Lessard, M., Michaux, L., Vyas, P., van Wering, E., Zwaan, C.M., Golub, T.R. & Orkin, S.H. (2006) Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. *Proceedings of the National Academy of Sciences U S A*, **103**, 3339-3344.
- Bowie, M.B., Kent, D.G., Dykstra, B., McKnight, K.D., McCaffrey, L., Hoodless, P.A. & Eaves, C.J. (2007) Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. *Proceedings of the National Academy of Sciences U S A*, **104**, 5878-5882.
- Boyes, J., Byfield, P., Nakatani, Y. & Ogryzko, V. (1998) Regulation of activity of the transcription factor GATA-1 by acetylation. *Nature*, **396**, 594-598.
- Bradley, A., Evans, M., Kaufman, M.H. & Robertson, E. (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, **309**, 255-256.
- Bruce, W.R. & Van Der Gaag, H. (1963) A Quantitative Assay for the Number of Murine Lymphoma Cells Capable of Proliferation in Vivo. *Nature*, **199**, 79-80.
- Cai, Z., de Bruijn, M., Ma, X., Dortland, B., Luteijn, T., Downing, R.J. & Dzierzak, E. (2000) Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity*, **13**, 423-431.
- Calligaris, R., Bottardi, S., Cogoi, S., Apezteguia, I. & Santoro, C. (1995) Alternative translation initiation site usage results in two functionally distinct forms of the GATA-1 transcription factor. *Proceedings of the National Academy of Sciences U S A*, **92**, 11598-11602.
- Calvi, L.M., Adams, G.B., Weibrecht, K.W., Weber, J.M. & Olson, D.P. (2003) Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature*, **425**, 841.
- Capecchi, M.R. (2001) Generating mice with targeted mutations. *Nature Medicine*, **7**, 1086-1090.
- Carpenter, E., Valverde-Garduno, V., Sternberg, A., Mitchell, C., Roberts, I., Vyas, P. & Vora, A. (2005) GATA1 mutation and trisomy 21 are required only in haematopoietic cells for development of transient myeloproliferative disorder. *British Journal of Haematology*, **128**, 548-551.
- Chen, C.Z., Li, L., Lodish, H.F. & Bartel, D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83-86.
- Chou, S.T., Opalinska, J.B., Yao, Y., Fernandes, M.A., Kalota, A., Brooks, J.S., Choi, J.K., Gewirtz, A.M., Danet-Desnoyers, G.A., Nemiroff, R.L. & Weiss, M.J. (2008) Trisomy 21 enhances human fetal erythro-megakaryocytic development. *Blood*.
- Collavin, L., Gostissa, M., Avolio, F., Secco, P., Ronchi, A., Santoro, C. & Del Sal, G. (2004) Modification of the erythroid transcription factor GATA-1 by SUMO-1. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 8870-8875.

- Corbel, C., Salan, J., Belo-Diabangouaya, P. & Diterlen-Livre, F. (2007) Hematopoietic potential of the pre-fusion allantois. *Developmental Biology*, **301**, 478.
- Corsetti, M.T. & Calabi, F. (1997) Lineage- and stage-specific expression of runt box polypeptides in primitive and definitive hematopoiesis. *Blood*, **89**, 2359-2368.
- Cotton, R.G. (1997) Slowly but surely towards better scanning for mutations. *Trends in Genetics*, **13**, 43-46.
- Cozzio, A., Passegue, E., Ayton, P.M., Karsunky, H., Cleary, M.L. & Weissman, I.L. (2003) Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. *Genes and Development*, **17**, 3029-3035.
- Crispino, J.D. (2005) GATA1 in normal and malignant hematopoiesis. *Seminars in Cell and Developmental Biology*, **16**, 137-147.
- Crispino, J.D., Lodish, M.B., MacKay, J.P. & Orkin, S.H. (1999) Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Molecular Cell*, **3**, 219-228.
- Crossley, M., Merika, M. & Orkin, S.H. (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Molecular and Cellular Biology*, **15**, 2448-2456.
- Crossley, M. & Orkin, S.H. (1994) Phosphorylation of the erythroid transcription factor GATA-1. *Journal of Biological Chemistry*, **269**, 16589-16596.
- Cuenca, G.M. & Ren, R. (2001) Cooperation of BCR-ABL and AML1/MDS1/EV11 in blocking myeloid differentiation and rapid induction of an acute myelogenous leukemia. *Oncogene*, **20**, 8236-8248.
- Cumano, A., Dieterlen-Livre, F. & Godin, I. (1996) Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura. *Cell*, **86**, 907.
- Cumano, A. & Godin, I. (2007) Ontogeny of the Hematopoietic System. *Annual Review of Immunology*, **25**, 745-785.
- Dash, A.B., Williams, I.R., Kutok, J.L., Tomasson, M.H., Anastasiadou, E., Lindahl, K., Li, S., Van Etten, R.A., Borrow, J., Housman, D., Druker, B. & Gilliland, D.G. (2002) A murine model of CML blast crisis induced by cooperation between BCR/ABL and NUP98/HOXA9. *Proceedings of the National Academy of Sciences U S A*, **99**, 7622-7627.
- Davis, R.L., Weintraub, H. & Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, **51**, 987-1000.
- De Vita, S., Devoy, A., Groet, J., Kruslin, B., Kuzmic-Prusac, I. & Nizetic, D. (2008) Megakaryocyte hyperproliferation without GATA1 mutation in foetal liver of a case of Down syndrome with hydrops foetalis. *British Journal of Haematology*.



- De Vita, S., Mulligan, C., McElwaine, S., Dagna-Bricarelli, F., Spinelli, M., Basso, G., Nizetic, D. & Groet, J. (2007) Loss-of-function JAK3 mutations in TMD and AMKL of Down syndrome. *British Journal of Haematology*, **137**, 337-341.
- DeKoter, R.P. & Singh, H. (2000) Regulation of B Lymphocyte and Macrophage Development by Graded Expression of PU.1. *Science*, **288**, 1439.
- Deutsch, E., Dugray, A., AbdulKarim, B., Marangoni, E., Maggiorella, L., Vaganay, S., M'Kacher, R., Rasy, S.D., Eschwege, F., Vainchenker, W., Turhan, A.G. & Bourhis, J. (2001) BCR-ABL down-regulates the DNA repair protein DNA-PKcs. *Blood*, **97**, 2084-2090.
- Deutsch, V.R. & Tomer, A. (2006) Megakaryocyte development and platelet production. *British Journal of Haematology*, **134**, 453-466.
- Dieterlen-Livre, F. (1975) On the origin of haemopoietic stem cells in the avian embryo: an experimental approach. *Journal of Embryology and Experimental Morphology*, **33**, 607.
- Dore, L.C., Amigo, J.D., Dos Santos, C.O., Zhang, Z., Gai, X., Tobias, J.W., Yu, D., Klein, A.M., Dorman, C., Wu, W., Hardison, R.C., Paw, B.H. & Weiss, M.J. (2008) A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proceedings of the National Academy of Sciences U S A*, **105**, 3333-3338.
- Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S. & de Laat, W. (2004) The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes and Development*, **18**, 2485-2490.
- Druker, B.J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G.M., Fanning, S., Zimmermann, J. & Lydon, N.B. (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Medicine*, **2**, 561-566.
- Eguchi, M., Eguchi-Ishimae, M., Knight, D., Kearney, L., Slany, R. & Greaves, M. (2006) MLL chimeric protein activation renders cells vulnerable to chromosomal damage: an explanation for the very short latency of infant leukemia. *Genes Chromosomes Cancer*, **45**, 754-760.
- Elagib, K.E. & Goldfarb, A.N. (2007) Regulation of RUNX1 transcriptional function by GATA-1. *Critical Reviews in Eukaryotic Gene Expression*, **17**, 271-280.
- Elagib, K.E., Racke, F.K., Mogass, M., Khetawat, R., Delehanty, L.L. & Goldfarb, A.N. (2003) RUNX1 and GATA-1 coexpression and cooperation in megakaryocytic differentiation. *Blood*, **101**, 4333-4341.
- Elrick, L.J., Jorgensen, H.G., Mountford, J.C. & Holyoake, T.L. (2005) Punish the parent not the progeny. *Blood*, **105**, 1862-1866.
- Emambokus, N.R. & Frampton, J. (2003) The Glycoprotein IIb Molecule Is Expressed on Early Murine Hematopoietic Progenitors and Regulates Their Numbers in Sites of Hematopoiesis. *Immunity*, **19**, 33-45.

- Eto, K., Murphy, R., Kerrigan, S.W., Berton, A., Stuhlmann, H., Nakano, T., Leavitt, A.D. & Shattil, S.J. (2002) Megakaryocytes derived from embryonic stem cells implicate CalDAG-GEFI in integrin signaling. *Proceedings of the National Academy of Sciences U S A*, **99**, 12819-12824.
- Evans, T., Reitman, M. & Felsenfeld, G. (1988) An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proceedings of the National Academy of Sciences U S A*, **85**, 5976-5980.
- Ferkowicz, M.J., Starr, M., Xie, X., Li, W. & Johnson, S.A. (2003) CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development*, **130**, 4393.
- Ferreira, R., Ohneda, K., Yamamoto, M. & Philipsen, S. (2005) GATA1 function, a paradigm for transcription factors in hematopoiesis. *Molecular and Cellular Biology*, **25**, 1215-1227.
- Ford, A.M., Bennett, C.A., Healy, L.E., Towatari, M., Greaves, M.F. & Enver, T. (1996) Regulation of the myeloperoxidase enhancer binding proteins Pu1, C-EBP alpha, -beta, and -delta during granulocyte-lineage specification. *Proceedings of the National Academy of Sciences USA*, **93**, 10838-10843.
- Forsberg, E.C., Serwold, T., Kogan, S., Weissman, I.L. & Passegue, E. (2006) New evidence supporting megakaryocyte-erythrocyte potential of flk2/flt3+ multipotent hematopoietic progenitors. *Cell*, **126**, 415-426.
- Freson, K., Devriendt, K., Matthijs, G., Van Hoof, A., De Vos, R., Thys, C., Minner, K., Hoylaerts, M.F., Vermyn, J. & Van Geet, C. (2001) Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. *Blood*, **98**, 85-92.
- Freson, K., Matthijs, G., Thys, C., Marien, P., Hoylaerts, M.F., Vermyn, J. & Van Geet, C. (2002) Different substitutions at residue D218 of the X-linked transcription factor GATA1 lead to altered clinical severity of macrothrombocytopenia and anemia and are associated with variable skewed X inactivation. *Human Molecular Genetics*, **11**, 147-152.
- Fujimoto, T.T., Kohata, S., Suzuki, H., Miyazaki, H. & Fujimura, K. (2003) Production of functional platelets by differentiated embryonic stem (ES) cells in vitro. *Blood*, **102**, 4044-4051.
- Fujita, Y., Nishimura, M., Taniwaki, M., Abe, T. & Okuda, T. (2001) Identification of an alternatively spliced form of the mouse AML1/RUNX1 gene transcript AML1c and its expression in early hematopoietic development. *Biochemical and Biophysical Research Communications*, **281**, 1248-1255.
- Fujiwara, Y., Browne, C.P., Cunliffe, K., Goff, S.C. & Orkin, S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proceedings of the National Academy of Sciences U S A*, **93**, 12355-12358.

- Gao, K., Masuda, A., Matsuura, T. & Ohno, K. (2008) Human branch point consensus sequence is yUnAy. *Nucleic Acids Research*, **36**, 2257-2267.
- Garzon, R. & Croce, C.M. (2008) MicroRNAs in normal and malignant hematopoiesis. *Current Opinion in Hematology*, **15**, 352-358.
- Garzon, R., Pichiorri, F., Palumbo, T., Iuliano, R., Cimmino, A., Aqeilan, R., Volinia, S., Bhatt, D., Alder, H., Marcucci, G., Calin, G.A., Liu, C.-G., Bloomfield, C.D., Andreeff, M. & Croce, C.M. (2006) MicroRNA fingerprints during human megakaryocytopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 5078-5083.
- Ge, Y., Dombkowski, A.A., LaFiura, K.M., Tatman, D., Yedidi, R.S., Stout, M.L., Buck, S.A., Massey, G., Becton, D.L., Weinstein, H.J., Ravindranath, Y., Matherly, L.H. & Taub, J.W. (2006) Differential gene expression, GATA1 target genes, and the chemotherapy sensitivity of Down syndrome megakaryocytic leukemia. *Blood*, **107**, 1570-1581.
- Gebauer, F. & Hentze, M.W. (2004) Molecular mechanisms of translational control. *Nature Reviews Molecular Cell Biology*, **5**, 827-835.
- Gekas, C., Dieterlen-Livre, F., Orkin, S.H. & Mikkola, H.K. (2005) The placenta is a niche for hematopoietic stem cells. *Developmental Cell*, **8**, 365.
- Ghinassi, B., Sanchez, M., Martelli, F., Amabile, G., Vannucchi, A.M., Migliaccio, G., Orkin, S.H. & Migliaccio, A.R. (2007) The hypomorphic Gata1low mutation alters the proliferation/differentiation potential of the common megakaryocytic-erythroid progenitor. *Blood*, **109**, 1460-1471.
- Ghirlando, R. & Trainor, C.D. (2003) Determinants of GATA-1 binding to DNA: the role of non-finger residues. *Journal of Biological Chemistry*, **278**, 45620-45628.
- Gilliland, D.G. (2001) Hematologic malignancies. *Current Opinion in Hematology*, **8**, 189-191.
- Godin, I.E., Garcia-Porrero, J.A., Coutinho, A., Dieterlen-Lievre, F. & Marcos, M.A.R. (1993) Para-aortic splanchnopleura from early mouse embryos contains B1a cell progenitors. *Nature*, **364**, 67-70.
- Goldman, L.A., Cutrone, E.C., Kotenko, S.V., Krause, C.D. & Langer, J.A. (1996) Modifications of vectors pEF-BOS, pcDNA1 and pcDNA3 result in improved convenience and expression. *Biotechniques*, **21**, 1013-1015.
- Grass, J.A., Boyer, M.E., Pal, S., Wu, J., Weiss, M.J. & Bresnick, E.H. (2003) GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proceedings of the National Academy of Sciences U S A*, **100**, 8811-8816.
- Greaves, M. (2005) In utero origins of childhood leukaemia. *Early Human Development*, **81**, 123-129.

Greaves, M. (2006) Infection, immune responses and the aetiology of childhood leukaemia. *Nature Reviews Cancer*, **6**, 193-203.

Greaves, M.F. (1986) Differentiation-linked leukemogenesis in lymphocytes. *Science*, **234**, 697-704.

Groet, J., McElwaine, S., Spinelli, M., Rinaldi, A., Burtscher, I., Mulligan, C., Mensah, A., Cavani, S., Dagna-Bricarelli, F., Basso, G., Cotter, F.E. & Nizetic, D. (2003) Acquired mutations in GATA1 in neonates with Down's syndrome with transient myeloid disorder. *Lancet*, **361**, 1617-1620.

Gurish, M.F. & Boyce, J.A. (2006) Mast cells: ontogeny, homing, and recruitment of a unique innate effector cell. *Journal of Allergy and Clinical Immunology*, **117**, 1285-1291.

Gutierrez, L., Lindeboom, F., Langeveld, A., Grosveld, F., Philipsen, S. & Whyatt, D. (2004) Homotypic signalling regulates Gata1 activity in the erythroblastic island. *Development*, **131**, 3183-3193.

Gutierrez, L., Nikolic, T., van Dijk, T.B., Hammad, H., Vos, N., Willart, M., Grosveld, F., Philipsen, S. & Lambrecht, B.N. (2007) Gata1 regulates dendritic-cell development and survival. *Blood*, **110**, 1933-1941.

Guyot, B., Valverde-Garduno, V., Porcher, C. & Vyas, P. (2004) Deletion of the major GATA1 enhancer HS 1 does not affect eosinophil GATA1 expression and eosinophil differentiation. *Blood*, **104**, 89-91.

Gwack, Y., Sharma, S., Nardone, J., Tanasa, B., Iuga, A., Srikanth, S., Okamura, H., Bolton, D., Feske, S., Hogan, P.G. & Rao, A. (2006) A genome-wide Drosophila RNAi screen identifies DYRK-family kinases as regulators of NFAT. *Nature*, **441**, 646-650.

Hama, A., Yagasaki, H., Takahashi, Y., Nishio, N., Muramatsu, H., Yoshida, N., Tanaka, M., Hidaka, H., Watanabe, N., Yoshimi, A., Matsumoto, K., Kudo, K., Kato, K., Horibe, K. & Kojima, S. (2008) Acute megakaryoblastic leukaemia (AMKL) in children: a comparison of AMKL with and without Down syndrome. *British Journal of Haematology*, **140**, 552-561.

Hannon, R., Evans, T., Felsenfeld, G. & Gould, H. (1991) Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proceedings of the National Academy of Sciences U S A*, **88**, 3004-3008.

Harigae, H., Takahashi, S., Suwabe, N., Ohtsu, H., Gu, L., Yang, Z., Tsai, F.Y., Kitamura, Y., Engel, J.D. & Yamamoto, M. (1998) Differential roles of GATA-1 and GATA-2 in growth and differentiation of mast cells. *Genes to Cells*, **3**, 39-50.

Harigae, H., Xu, G., Sugawara, T., Ishikawa, I., Toki, T. & Ito, E. (2004) The GATA1 mutation in an adult patient with acute megakaryoblastic leukemia not accompanying Down syndrome. *Blood*, **103**, 3242-3243.

Hasle, H., Abrahamsson, J., Arola, M., Karow, A., O'Marcaigh, A., Reinhardt, D., Webb, D.K., van Wering, E., Zeller, B., Zwaan, C.M. & Vyas, P. (2008) Myeloid leukemia in

children 4 years or older with Down syndrome often lacks GATA1 mutation and cytogenetics and risk of relapse are more akin to sporadic AML. *Leukemia*, **22**, 1428-1430.

Hasle, H., Clemmensen, I.H. & Mikkelsen, M. (2000) Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet*, **355**, 165-169.

Hasty, P., Rivera-Perez, J. & Bradley, A. (1991) The length of homology required for gene targeting in embryonic stem cells. *Molecular and Cellular Biology*, **11**, 5586-5591.

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Groner, Y., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Schibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordsiek, G., Hornischer, K., Brant, P., Scharfe, M., Schon, O., Desario, A., Reichelt, J., Kauer, G., Blocker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Haaf, T., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. & Yaspo, M.L. (2000) The DNA sequence of human chromosome 21. *Nature*, **405**, 311-319.

He, L. & Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, **5**, 522-531.

Heald, B., Hilden, J.M., Zbuk, K., Norton, A., Vyas, P., Theil, K.S. & Eng, C. (2007) Severe TMD/AMKL with GATA1 mutation in a stillborn fetus with Down syndrome. *Nature Clinical Practice Oncology*, **4**, 433-438.

Hernandez-Hernandez, A., Ray, P., Litos, G., Ciro, M., Ottolenghi, S., Beug, H. & Boyes, J. (2006) Acetylation and MAPK phosphorylation cooperate to regulate the degradation of active GATA-1. *European Molecular Biology Organisation Journal*, **25**, 3264-3274.

Herzog, E.L., Chai, L. & Krause, D.S. (2003) Plasticity of marrow-derived stem cells. *Blood*, **102**, 3483-3493.

Higuchi, M., O'Brien, D., Kumaravelu, P., Lenny, N., Yeoh, E.J. & Downing, J.R. (2002) Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell*, **1**, 63-74.

Hitzler, J.K., Cheung, J., Li, Y., Scherer, S.W. & Zipursky, A. (2003) GATA1 mutations in transient leukemia and acute megakaryoblastic leukemia of Down syndrome. *Blood*, **101**, 4301-4304.

Hitzler, J.K. & Zipursky, A. (2005) Origins of leukaemia in children with Down syndrome. *Nature Reviews Cancer*, **5**, 11-20.

Hoffbrand, A.V., Catovsky, D. & Tuddenham, E.G.D. (2005) *Postgraduate haematology*. Blackwell, Oxford.

- Holland, L.M., Lima, C.S., Cunha, A.F., Albuquerque, D.M., Vassallo, J., Ozelo, M.C., Joazeiro, P.P., Saad, S.T. & Costa, F.F. (2006) An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nature Genetics*, **38**, 807-812.
- Holmes, D.K., Bates, N., Murray, M., Ladusans, E.J., Morabito, A., Bolton-Maggs, P.H., Johnston, T.A., Walkenshaw, S., Wynn, R.F. & Bellantuono, I. (2006) Hematopoietic progenitor cell deficiency in fetuses and children affected by Down's syndrome. *Experimental Hematology*, **34**, 1611-1615.
- Holt, S.E., Brown, E.J. & Zipursky, A. (2002) Telomerase and the benign and malignant megakaryoblastic leukemias of Down syndrome. *Journal of Pediatric Hematology and Oncology*, **24**, 14-17.
- Hong, D., Gupta, R., Ancliff, P., Atzberger, A., Brown, J., Soneji, S., Green, J., Colman, S., Piacibello, W., Buckle, V., Tsuzuki, S., Greaves, M. & Enver, T. (2008) Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science*, **319**, 336-339.
- Hooper, M., Hardy, K., Handyside, A., Hunter, S. & Monk, M. (1987) HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature*, **326**, 292-295.
- Hu, M., Krause, D., Greaves, M., Sharkis, S., Dexter, M., Heyworth, C. & Enver, T. (1997) Multilineage gene expression precedes commitment in the hemopoietic system. *Genes and Development*, **11**, 774-785.
- Huber, T.L., Kouskoff, V., Fehling, H.J., Palis, J. & Keller, G. (2004) Haemangioblast commitment is initiated in the primitive streak of the mouse embryo. *Nature*, **432**, 625.
- Hung, H.L., Lau, J., Kim, A.Y., Weiss, M.J. & Blobel, G.A. (1999) CREB-Binding protein acetylates hematopoietic transcription factor GATA-1 at functionally important sites. *Molecular and Cellular Biology*, **19**, 3496-3505.
- Ichikawa, M., Asai, T., Chiba, S., Kurokawa, M. & Ogawa, S. (2004) Runx1/AML-1 ranks as a master regulator of adult hematopoiesis. *Cell Cycle*, **3**, 722-724.
- Ikonomi, P., Rivera, C.E., Riordan, M., Washington, G., Schechter, A.N. & Noguchi, C.T. (2000) Overexpression of GATA-2 inhibits erythroid and promotes megakaryocyte differentiation. *Experimental Hematology*, **28**, 1423-1431.
- Ito, E., Kasai, M., Hayashi, Y., Toki, T., Arai, K., Yokoyama, S., Kato, K., Tachibana, N., Yamamoto, M. & Yokoyama, M. (1995) Expression of erythroid-specific genes in acute megakaryoblastic leukaemia and transient myeloproliferative disorder in Down's syndrome. *British Journal of Haematology*, **90**, 607-614.
- Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J.D. & Yamamoto, M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature*, **362**, 466-468.

- Jackson, M., Baird, J.W., Cambray, N., Ansell, J.D., Forrester, L.M. & Graham, G.J. (2002) Cloning and Characterization of Ehox, a Novel Homeobox Gene Essential for Embryonic Stem Cell Differentiation. *Journal of Biological Chemistry*, **277**, 38683-38692.
- Jaenisch, R. & Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, **33 Suppl**, 245-254.
- Jaffe, E.S. (2001) *Pathology and genetics of tumours of haematopoietic and lymphoid tissues*. IARC Press ; Oxford : Oxford University Press [distributor], Lyon.
- Johansson, B.M. & Wiles, M.V. (1995) Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development. *Molecular and Cellular Biology*, **15**, 141.
- Johnnidis, J.B., Harris, M.H., Wheeler, R.T., Stehling-Sun, S., Lam, M.H., Kirak, O., Brummelkamp, T.R., Fleming, M.D. & Camargo, F.D. (2008) Regulation of progenitor cell proliferation and granulocyte function by microRNA-223. *Nature*, **451**, 1125-1129.
- Jung, Y.J., Chae, H.C., Seoh, J.Y., Ryu, K.H., Park, H.K., Kim, Y.J. & Woo, S.Y. (2007) Pim-1 induced polyploidy but did not affect megakaryocytic differentiation of K562 cells and CD34+ cells from cord blood. *European Journal of Haematology*, **78**, 131-138.
- Kadri, Z., Maouche-Chretien, L., Rooke, H.M., Orkin, S.H., Romeo, P.H., Mayeux, P., Leboulch, P. & Chretien, S. (2005) Phosphatidylinositol 3-kinase/Akt induced by erythropoietin renders the erythroid differentiation factor GATA-1 competent for TIMP-1 gene transactivation. *Molecular and Cellular Biology*, **25**, 7412-7422.
- Keller, G., Kennedy, M., Papayannopoulou, T. & Wiles, M.V. (1993) Hematopoietic commitment during embryonic stem cell differentiation in culture. *Molecular and Cellular Biology*, **13**, 473-486.
- Kelly, L.M., Englmeier, U., Lafon, I., Sieweke, M.H. & Graf, T. (2000) MafB is an inducer of monocytic differentiation. *European Molecular Biology Organisation Journal*, **19**, 1987-1997.
- Kiel, M.J., Yilmaz, O.H., Iwashita, T., Terhorst, C. & Morrison, S.J. (2005) SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell*, **121**, 1109.
- Kingsley, P.D., Malik, J., Emerson, R.L., Bushnell, T.P., McGrath, K.E., Bloedorn, L.A., Bulger, M. & Palis, J. (2006) "Maturation" globin switching in primary primitive erythroid cells. *Blood*, **107**, 1665-1672.
- Kirsammer, G., Jilani, S., Liu, H., Davis, E., Gurbuxani, S., Le Beau, M.M. & Crispino, J.D. (2008) Highly penetrant myeloproliferative disease in the Ts65Dn mouse model of Down syndrome. *Blood*, **111**, 767-775.
- Kivivuori, S.M., Rajantie, J. & Siimes, M.A. (1996) Peripheral blood cell counts in infants with Down's syndrome. *Clinical Genetics*, **49**, 15-19.

- Kiyoi, H., Yamaji, S., Kojima, S. & Naoe, T. (2007) JAK3 mutations occur in acute megakaryoblastic leukemia both in Down syndrome children and non-Down syndrome adults. *Leukemia*, **21**, 574-576.
- Klusmann, J.H., Creutzig, U., Zimmermann, M., Dworzak, M., Jorch, N., Langebrake, C., Pekrun, A., Macakova-Reinhardt, K. & Reinhardt, D. (2008) Treatment and prognostic impact of transient leukemia in neonates with Down syndrome. *Blood*, **111**, 2991-2998.
- Klusmann, J.H., Reinhardt, D., Hasle, H., Kaspers, G.J., Creutzig, U., Hahlen, K., van den Heuvel-Eibrink, M.M. & Zwaan, C.M. (2007) Janus kinase mutations in the development of acute megakaryoblastic leukemia in children with and without Down's syndrome. *Leukemia*, **21**, 1584-1587.
- Knudson, A.G. (2001) Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, **1**, 157-162.
- Knuesel, M., Wan, Y., Xiao, Z., Holinger, E., Lowe, N., Wang, W. & Liu, X. (2003) Identification of Novel Protein-Protein Interactions Using A Versatile Mammalian Tandem Affinity Purification Expression System. *Molecular and Cellular Proteomics*, **2**, 1225-1233.
- Kondo, M., Weissman, I.L. & Akashi, K. (1997) Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, **91**, 661-672.
- Kornblau, S.M., Womble, M., Qiu, Y.H., Jackson, C.E., Chen, W., Konopleva, M., Estey, E.H. & Andreeff, M. (2006) Simultaneous activation of multiple signal transduction pathways confers poor prognosis in acute myelogenous leukemia. *Blood*, **108**, 2358-2365.
- Kozak, M. (1986a) Bifunctional messenger RNAs in eukaryotes. *Cell*, **47**, 481-483.
- Kozak, M. (1986b) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283-292.
- Kuhl, C., Atzberger, A., Iborra, F., Nieswandt, B., Porcher, C. & Vyas, P. (2005) GATA1-mediated megakaryocyte differentiation and growth control can be uncoupled and mapped to different domains in GATA1. *Molecular and Cellular Biology*, **25**, 8592-8606.
- Kulesa, H., Frampton, J. & Graf, T. (1995) GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblats, and erythroblats. *Genes and Development*, **9**, 1250-1262.
- Kurahashi, H., Hara, J., Yumura-Yagi, K., Murayama, N., Inoue, M., Ishihara, S., Tawa, A., Okada, S. & Kawa-Ha, K. (1991) Monoclonal nature of transient abnormal myelopoiesis in Down's syndrome. *Blood*, **77**, 1161-1163.
- Lange, B.J., Kobrinsky, N., Barnard, D.R., Arthur, D.C., Buckley, J.D., Howells, W.B., Gold, S., Sanders, J., Neudorf, S., Smith, F.O. & Woods, W.G. (1998) Distinctive demography, biology, and outcome of acute myeloid leukemia and myelodysplastic syndrome in children with Down syndrome: Children's Cancer Group Studies 2861 and 2891. *Blood*, **91**, 608-615.



- le Viseur, C., Hotfilder, M., Bomken, S., Wilson, K., Rottgers, S., Schrauder, A., Rosemann, A., Irving, J., Stam, R.W., Shultz, L.D., Harbott, J., Jurgens, H., Schrappe, M., Pieters, R. & Vormoor, J. (2008) In childhood acute lymphoblastic leukemia, blasts at different stages of immunophenotypic maturation have stem cell properties. *Cancer Cell*, **14**, 47-58.
- Leder, A., Kuo, A., Shen, M.M. & Leder, P. (1992) In situ hybridization reveals co-expression of embryonic and adult alpha globin genes in the earliest murine erythrocyte progenitors. *Development*, **116**, 1041-1049.
- Letting, D.L., Rakowski, C., Weiss, M.J. & Blobel, G.A. (2003) Formation of a tissue-specific histone acetylation pattern by the hematopoietic transcription factor GATA-1. *Molecular and Cellular Biology*, **23**, 1334-1340.
- Li, C.-Y., Fang, F., Xu, W.-X., Xu, C.-w., Zhan, Y.-Q., Wang, Z.-D., Ding, Y.-l., Li, Y.-H., Sun, H.-B. & Yang, X.-M. (2008a) Suppression of EDAG gene expression by phorbol 12-myristate 13-acetate is mediated through down-regulation of GATA-1. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, **1779**, 606-615.
- Li, Y., Lu, J., Cohen, D. & Prochownik, E.V. (2008b) Transformation, genomic instability and senescence mediated by platelet/megakaryocyte glycoprotein Ibalpha. *Oncogene*, **27**, 1599-1609.
- Li, Z., Godinho, F.J., Klusmann, J.H., Garriga-Canut, M., Yu, C. & Orkin, S.H. (2005) Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nature Genetics*, **37**, 613-619.
- Lightfoot, J., Hitzler, J.K., Zipursky, A., Albert, M. & Macgregor, P.F. (2004) Distinct gene signatures of transient and acute megakaryoblastic leukemia in Down syndrome. *Leukemia*, **18**, 1617-1623.
- Lindeboom, F., Gillemans, N., Karis, A., Jaegle, M., Meijer, D., Grosveld, F. & Philipsen, S. (2003) A tissue-specific knockout reveals that Gata1 is not essential for Sertoli cell function in the mouse. *Nucleic Acids Research*, **31**, 5405-5412.
- Ling, K.W., Ottersbach, K., van Hamburg, J.P., Oziemlak, A., Tsai, F.Y., Orkin, S.H., Ploemacher, R., Hendriks, R.W. & Dzierzak, E. (2004) GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *Journal of Experimental Medicine*, **200**, 871-882.
- Liu, H.S., Jan, M.S., Chou, C.K., Chen, P.H. & Ke, N.J. (1999) Is green fluorescent protein toxic to the living cells? *Biochemical and Biophysical Research Communications*, **260**, 712-717.
- Longo, L., Bygrave, A., Grosveld, F.G. & Pandolfi, P.P. (1997) The chromosome make-up of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. *Transgenic Research*, **6**, 321-328.
- Lowry, J.A., Gamsjaeger, R., Thong, S.Y., Hung, W., Kwan, A.H., Broitman-Maduro, G., Matthews, J.M., Maduro, M. & Mackay, J.P. (2008) Structural analysis of med-1 reveals

unexpected diversity in the mechanism of DNA recognition by gata-type zinc finger domains. (2009) *Journal of Biological Chemistry*, **284**, 5827-35.

Lozzio, B.B. & Lozzio, C.B. (1979) Properties and usefulness of the original K-562 human myelogenous leukemia cell line. *Leukemia Research*, **3**, 363-370.

Lyons, S.E., Lawson, N.D., Lei, L., Bennett, P.E., Weinstein, B.M. & Liu, P.P. (2002) A nonsense mutation in zebrafish gata1 causes the bloodless phenotype in vlad tepes. *Proceedings of the National Academy of Sciences U S A*, **99**, 5454-5459.

Ma, Z., Morris, S.W., Valentine, V., Li, M., Herbrick, J.A., Cui, X., Bouman, D., Li, Y., Mehta, P.K., Nizetic, D., Kaneko, Y., Chan, G.C., Chan, L.C., Squire, J., Scherer, S.W. & Hitzler, J.K. (2001) Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. *Nature Genetics*, **28**, 220-221.

Magalhaes, I.Q., Splendore, A., Emerenciano, M., Figueiredo, A., Ferrari, I. & Pombo-de-Oliveira, M.S. (2006) GATA1 mutations in acute leukemia in children with Down syndrome. *Cancer Genetics and Cytogenetics*, **166**, 112-116.

Majewski, I.J., Metcalf, D., Mielke, L.A., Krebs, D.L., Ellis, S., Carpinelli, M.R., Mifsud, S., Di Rago, L., Corbin, J., Nicola, N.A., Hilton, D.J. & Alexander, W.S. (2006) A mutation in the translation initiation codon of Gata-1 disrupts megakaryocyte maturation and causes thrombocytopenia. *Proceedings of the National Academy of Sciences U S A*, **103**, 14146-14151.

Mansour, S.L., Thomas, K.R. & Capecchi, M.R. (1988) Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature*, **336**, 348-352.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., Buza-Vidas, N., Qian, H., Soneji, S., Enver, T., Sigvardsson, M. & Jacobsen, S.E.W. (2007) Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors. *Immunity*, **26**, 407-419.

Martin, D.I. & Orkin, S.H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes and Development*, **4**, 1886-1898.

Massey, G.V., Zipursky, A., Chang, M.N., Doyle, J.J., Nasim, S., Taub, J.W., Ravindranath, Y., Dahl, G. & Weinstein, H.J. (2006) A prospective study of the natural history of transient leukemia (TL) in neonates with Down syndrome (DS): Children's Oncology Group (COG) study POG-9481. *Blood*, **107**, 4606-4613.

Masuda, A., Hashimoto, K., Yokoi, T., Doi, T., Kodama, T., Kume, H., Ohno, K. & Matsuguchi, T. (2007) Essential Role of GATA Transcriptional Factors in the Activation of Mast Cells. *Journal of Immunology*, **178**, 360-368.

Mata, J., Marguerat, S. & Bahler, J. (2005) Post-transcriptional control of gene expression: a genome-wide perspective. *Trends in Biochemical Sciences*, **30**, 506-514.

- Matsumura, I., Tanaka, H., Kawasaki, A., Odajima, J., Daino, H., Hashimoto, K., Wakao, H., Nakajima, K., Kato, T., Miyazaki, H. & Kanakura, Y. (2000) Increased D-type cyclin expression together with decreased cdc2 activity confers megakaryocytic differentiation of a human thrombopoietin-dependent hematopoietic cell line. *Journal of Biological Chemistry*, **275**, 5553-5559.
- Matsuoka, S., Tsuji, K., Hisakawa, H., Xu, M. & Ebihara, Y. (2001) Generation of definitive hematopoietic stem cells from murine early yolk sac and paraaortic splanchnopleures by aorta-gonad-mesonephros region-derived stromal cells. *Blood*, **98**, 6.
- McDevitt, M.A., Fujiwara, Y., Shivdasani, R.A. & Orkin, S.H. (1997a) An upstream, DNase I hypersensitive region of the hematopoietic-expressed transcription factor GATA-1 gene confers developmental specificity in transgenic mice. *Proceedings of the National Academy of Sciences U S A*, **94**, 7976-7981.
- McDevitt, M.A., Shivdasani, R.A., Fujiwara, Y., Yang, H. & Orkin, S.H. (1997b) A "knockdown" mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proceedings of the National Academy of Sciences U S A*, **94**, 6781-6785.
- McElwaine, S., Mulligan, C., Groet, J., Spinelli, M., Rinaldi, A., Denyer, G., Mensah, A., Cavani, S., Baldo, C., Dagna-Bricarelli, F., Hann, I., Basso, G., Cotter, F.E. & Nizetic, D. (2004) Microarray transcript profiling distinguishes the transient from the acute type of megakaryoblastic leukaemia (M7) in Down's syndrome, revealing PRAME as a specific discriminating marker. *British Journal of Haematology*, **125**, 729-742.
- McKimmie, C.S., Fraser, A.R., Hansell, C., Gutierrez, L., Philipsen, S., Connell, L., Rot, A., Kurowska-Stolarska, M., Carreno, P., Pruenster, M., Chu, C.C., Lombardi, G., Halsey, C., McInnes, I.B., Liew, F.Y., Nibbs, R.J. & Graham, G.J. (2008) Hemopoietic cell expression of the chemokine decoy receptor D6 is dynamic and regulated by GATA1. *Journal of Immunology*, **181**, 3353-3363.
- Medvinsky, A. & Dzierzak, E. (1996) Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell*, **86**, 897.
- Mehaffey, M.G., Newton, A.L., Gandhi, M.J., Crossley, M. & Drachman, J.G. (2001) X-linked thrombocytopenia caused by a novel mutation of GATA-1. *Blood*, **98**, 2681-2688.
- Melo, J.V. & Barnes, D.J. (2007) Chronic myeloid leukaemia as a model of disease evolution in human cancer. *Nature Reviews Cancer*, **7**, 441-453.
- Mendes, S.C., Robin, C. & Dzierzak, E. (2005) Mesenchymal progenitor cells localize within hematopoietic sites throughout ontogeny. *Development*, **132**, 1127.
- Mercher, T., Wernig, G., Moore, S.A., Levine, R.L., Gu, T.L., Frohling, S., Cullen, D., Polakiewicz, R.D., Bernard, O.A., Boggon, T.J., Lee, B.H. & Gilliland, D.G. (2006) JAK2T875N is a novel activating mutation that results in myeloproliferative disease with features of megakaryoblastic leukemia in a murine bone marrow transplantation model. *Blood*, **108**, 2770-2779.

- Metcalf, D. (1970) Studies on colony formation in vitro by mouse bone marrow cells. II. Action of colony stimulating factor. *Journal of Cell Physiology*, **76**, 89-99.
- Metcalf, D. (1991) Lineage commitment of hemopoietic progenitor cells in developing blast cell colonies: influence of colony-stimulating factors. *Proceedings of the National Academy of Sciences U S A*, **88**, 11310-11314.
- Metcalf, D. (1999) Cellular hematopoiesis in the twentieth century. *Seminars in Hematology*, **36**, 5-12.
- Meyers, E.N., Lewandoski, M. & Martin, G.R. (1998) An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nature Genetics*, **18**, 136-141.
- Migliaccio, A.R., Rana, R.A., Sanchez, M., Lorenzini, R., Centurione, L., Bianchi, L., Vannucchi, A.M., Migliaccio, G. & Orkin, S.H. (2003) GATA-1 as a regulator of mast cell differentiation revealed by the phenotype of the GATA-1low mouse mutant. *Journal of Experimental Medicine*, **197**, 281-296.
- Mikkola, H.K., Fujiwara, Y., Schlaeger, T.M., Traver, D. & Orkin, S.H. (2003) Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. *Blood*, **101**, 508.
- Misteli, T. & Spector, D.L. (1997) Applications of the green fluorescent protein in cell biology and biotechnology. *Nature Biotechnology*, **15**, 961-964.
- Mitjavila-Garcia, M.T., Cailleret, M., Godin, I., Nogueira, M.M., Cohen-Solal, K., Schiavon, V., Lecluse, Y., Le Pesteur, F., Lagrue, A.H. & Vainchenker, W. (2002) Expression of CD41 on hematopoietic progenitors derived from embryonic hematopoietic cells. *Development*, **129**, 2003-2013.
- Miyauchi, J., Ito, Y., Kawano, T., Tsunematsu, Y. & Shimizu, K. (1992) Unusual diffuse liver fibrosis accompanying transient myeloproliferative disorder in Down's syndrome: a report of four autopsy cases and proposal of a hypothesis. *Blood*, **80**, 1521-1527.
- Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. & Hayakawa, T. (2000) IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Molecular Therapy*, **1**, 376-382.
- Mizushima, S. & Nagata, S. (1990) pEF-BOS, a powerful mammalian expression vector. *Nucleic Acids Research*, **18**, 5322.
- Mori, H., Colman, S.M., Xiao, Z., Ford, A.M., Healy, L.E., Donaldson, C., Hows, J.M., Navarrete, C. & Greaves, M. (2002) Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Sciences U S A*, **99**, 8242-8247.
- Morle, F., Laverriere, A.C. & Godet, J. (1992) Globin genes are actively transcribed in the human megakaryoblastic leukemia cell line MEG-01. *Blood*, **79**, 3094-3096.

- Mountford, P.S. & Smith, A.G. (1995) Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis. *Trends in Genetics*, **11**, 179-184.
- Mroske, C., Plant, M.H., Franks, D.J. & Laneuville, O. (2000) Characterization of prostaglandin endoperoxide H synthase-1 enzyme expression during differentiation of the megakaryocytic cell line MEG-01. *Experimental Hematology*, **28**, 411-421.
- Muntean, A.G. & Crispino, J.D. (2005) Differential requirements for the activation domain and FOG-interaction surface of GATA-1 in megakaryocyte gene expression and development. *Blood*, **106**, 1223-1231.
- Murakami, T., Ohyashiki, K., Ohyashiki, J.H., Kawakubo, K., Tauchi, T., Nakazawa, S., Kimura, N. & Toyama, K. (1994) Cytogenetic and immunogenotypic alterations of blast crisis cells in chronic myelogenous leukemia independently linked to immunophenotypic expression. *Cancer Genetics and Cytogenetics*, **72**, 48-54.
- Nagaraj, S.H., Gasser, R.B. & Ranganathan, S. (2007) A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics*, **8**, 6-21.
- Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H. & Kishimoto, T. (1996) Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature*, **382**, 635-638.
- Nagata, K.I., Okano, Y. & Nozawa, Y. (1996) Protein kinase C isozymes in human megakaryoblastic leukemia cell line, MEG-01: possible involvement of the isozymes in the differentiation process of MEG-01 cells. *British Journal of Haematology*, **93**, 762-771.
- Nagy, A. (2003) *Manipulating the mouse embryo : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Nagy, A., Moens, C., Ivanyi, E., Pawling, J., Gertsenstein, M., Hadjantonakis, A.K., Pirity, M. & Rossant, J. (1998) Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. *Current Biology*, **8**, 661-664.
- Naito, M., Umeda, S., Yamamoto, T., Moriyama, H., Umezu, H., Hasegawa, G., Usuda, H., Shultz, L.D. & Takahashi, K. (1996) Development, differentiation, and phenotypic heterogeneity of murine tissue macrophages. *Journal of Leukocyte Biology*, **59**, 133-138.
- Nakao, M., Yokota, S., Iwai, T., Kaneko, H., Horiike, S., Kashima, K., Sonoda, Y., Fujimoto, T. & Misawa, S. (1996) Internal tandem duplication of the *flt3* gene found in acute myeloid leukemia. *Leukemia*, **10**, 1911-1918.
- Namciu, S., Lieberman, M.A. & Stavnezer, E. (1994) Induction of the *c-ski* proto-oncogene by phorbol ester correlates with induction of megakaryocyte differentiation. *Oncogene*, **9**, 1407-1416.
- Nerlov, C., Querfurth, E., Kulesa, H. & Graf, T. (2000) GATA-1 interacts with the myeloid PU.1 transcription factor and represses PU.1-dependent transcription. *Blood*, **95**, 2543-2551.

- Nichols, K.E., Crispino, J.D., Poncz, M., White, J.G., Orkin, S.H., Maris, J.M. & Weiss, M.J. (2000) Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nature Genetics*, **24**, 266-270.
- Nicolis, S., Bertini, C., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B. & Ottolenghi, S. (1991) An erythroid specific enhancer upstream to the gene encoding the cell-type specific transcription factor GATA-1. *Nucleic Acids Research*, **19**, 5285-5291.
- Nishimura, S., Takahashi, S., Kuroha, T., Suwabe, N., Nagasawa, T., Trainor, C. & Yamamoto, M. (2000) A GATA box in the GATA-1 gene hematopoietic enhancer is a critical element in the network of GATA factors and sites that regulate this gene. *Molecular and Cellular Biology*, **20**, 713-723.
- Norton, A., Fisher, C., Liu, H., Wen, Q., Mundschau, G., Fuster, J.L., Hasle, H., Zeller, B., Webb, D.K., O'Marcaigh, A., Sorrell, A., Hilden, J., Gamis, A., Crispino, J.D. & Vyas, P. (2007) Analysis of JAK3, JAK2, and C-MPL mutations in transient myeloproliferative disorder and myeloid leukemia of Down syndrome blasts in children with Down syndrome. *Blood*, **110**, 1077-1079.
- O'Doherty, A., Ruf, S., Mulligan, C., Hildreth, V., Errington, M.L., Cooke, S., Sesay, A., Modino, S., Vanes, L., Hernandez, D., Linehan, J.M., Sharpe, P.T., Brandner, S., Bliss, T.V., Henderson, D.J., Nizetic, D., Tybulewicz, V.L. & Fisher, E.M. (2005) An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science*, **309**, 2033-2037.
- Ogura, M., Morishima, Y., Ohno, R., Kato, Y., Hirabayashi, N., Nagura, H. & Saito, H. (1985) Establishment of a novel human megakaryoblastic leukemia cell line, MEG-01, with positive Philadelphia chromosome. *Blood*, **66**, 1384-1392.
- Ogura, M., Morishima, Y., Okumura, M., Hotta, T., Takamoto, S., Ohno, R., Hirabayashi, N., Nagura, H. & Saito, H. (1988) Functional and morphological differentiation induction of a human megakaryoblastic leukemia cell line (MEG-01s) by phorbol diesters. *Blood*, **72**, 49-60.
- Okuda, T., van Deursen, J., Hiebert, S.W., Grosveld, G. & Downing, J.R. (1996) AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell*, **84**, 321-330.
- Onodera, K., Takahashi, S., Nishimura, S., Ohta, J., Motohashi, H., Yomogida, K., Hayashi, N., Engel, J.D. & Yamamoto, M. (1997) GATA-1 transcription is controlled by distinct regulatory mechanisms during primitive and definitive erythropoiesis. *Proceedings of the National Academy of Sciences U S A*, **94**, 4487-4492.
- Orkin, S.H. (2000) Diversification of haematopoietic stem cells to specific lineages. *Nature Reviews Genetics*, **1**, 57-64.
- Ottersbach, K. & Dzierzak, E. (2005) The murine placenta contains hematopoietic stem cells within the vascular labyrinth region. *Developmental Cell*, **8**, 377.
- Owen, C.J., Toze, C.L., Kooschin, A., Forrest, D.L., Smith, C.A., Stevens, J.M., Jackson, S.C., Poon, M.C., Sinclair, G.D., Leber, B., Johnson, P.R., Macheta, A., Yin, J.A., Barnett,

- M.J., Lister, T.A. & Fitzgibbon, J. (2008) Five new pedigrees with inherited RUNX1 mutations causing familial platelet disorder with propensity to myeloid malignancy (FPD/AML). *Blood*, **112**, 4639-45.
- Pabst, T., Stillner, E., Neuberg, D., Nimer, S., Willman, C.L., List, A.F., Melo, J.V., Tenen, D.G. & Mueller, B.U. (2006) Mutations of the myeloid transcription factor CEBPA are not associated with the blast crisis of chronic myeloid leukaemia. *British Journal of Haematology*, **133**, 400-402.
- Pearson-White, S., Deacon, D., Crittenden, R., Brady, G., Iscove, N. & Quesenberry, P.J. (1995) The ski/sno protooncogene family in hematopoietic development. *Blood*, **86**, 2146-2155.
- Persons, D.A., Allay, J.A., Allay, E.R., Ashmun, R.A., Orlic, D., Jane, S.M., Cunningham, J.M. & Nienhuis, A.W. (1999) Enforced expression of the GATA-2 transcription factor blocks normal hematopoiesis. *Blood*, **93**, 488-499.
- Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H. & Costantini, F. (1995) Development of hematopoietic cells lacking transcription factor GATA-1. *Development*, **121**, 163-172.
- Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H. & Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257-260.
- Pikaart, M.J., Recillas-Targa, F. & Felsenfeld, G. (1998) Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes and Development*, **12**, 2852-2862.
- Pina, C. & Enver, T. (2007) Differential contributions of haematopoietic stem cells to foetal and adult haematopoiesis: insights from functional analysis of transcriptional regulators. *Oncogene*, **26**, 6750-6765.
- Pine, S.R., Guo, Q., Yin, C., Jayabose, S., Druschel, C.M. & Sandoval, C. (2007) Incidence and clinical implications of GATA1 mutations in newborns with Down syndrome. *Blood*, **110**, 2128-2131.
- Qian, H., Georges-Labouesse, E., Nystrom, A., Domogatskaya, A., Tryggvason, K., Jacobsen, S.E.W. & Ekblom, M. (2007) Distinct roles of integrins  $\alpha 6$  and  $\alpha 4$  in homing of fetal liver hematopoietic stem and progenitor cells. *Blood*, **110**, 2399-2407.
- Rainis, L., Bercovich, D., Strehl, S., Teigler-Schlegel, A., Stark, B., Trka, J., Amariglio, N., Biondi, A., Muler, I., Rechavi, G., Kempinski, H., Haas, O.A. & Izraeli, S. (2003) Mutations in exon 2 of GATA1 are early events in megakaryocytic malignancies associated with trisomy 21. *Blood*, **102**, 981-986.
- Rainis, L., Toki, T., Pimanda, J.E., Rosenthal, E., Machol, K., Strehl, S., Gottgens, B., Ito, E. & Izraeli, S. (2005) The proto-oncogene ERG in megakaryoblastic leukemias. *Cancer Research*, **65**, 7596-7602.

- Ray, S., Lu, Y., Kaufmann, S.H., Gustafson, W.C., Karp, J.E., Boldogh, I., Fields, A.P. & Brasier, A.R. (2004) Genomic mechanisms of p210BCR-ABL signaling: induction of heat shock protein 70 through the GATA response element confers resistance to paclitaxel-induced apoptosis. *Journal of Biological Chemistry*, **279**, 35604-35615.
- Recillas-Targa, F. (2006) Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. *Molecular Biotechnology*, **34**, 337-354.
- Rekhtman, N., Radparvar, F., Evans, T. & Skoultschi, A.I. (1999) Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells. *Genes and Development*, **13**, 1398-1411.
- Reya, T., Morrison, S.J., Clarke, M.F. & Weissman, I.L. (2001) Stem cells, cancer, and cancer stem cells. *Nature*, **414**, 105-111.
- Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. & Bartel, D.P. (2002) Prediction of plant microRNA targets. *Cell*, **110**, 513-520.
- Ribeil, J.A., Zermati, Y., Vandekerckhove, J., Cathelin, S., Kersual, J., Dussiot, M., Coulon, S., Moura, I.C., Zeuner, A., Kirkegaard-Sorensen, T., Varet, B., Solary, E., Garrido, C. & Hermine, O. (2007) Hsp70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. *Nature*, **445**, 102-105.
- Robertson, M., De Jong, G. & Mansvelt, E. (2003) Prenatal diagnosis of congenital leukemia in a fetus at 25 weeks' gestation with Down syndrome: case report and review of the literature. *Ultrasound in Obstetrics and Gynecology*, **21**, 486-489.
- Rodriguez, M.H., Enjolras, N., Plantier, J.L., Rea, M., Leboeuf, M., Uzan, G., Bordet, J.C. & Negrier, C. (2002) Expression of coagulation factor IX in a haematopoietic cell line. *Thrombosis and Haemostasis*, **87**, 366-373.
- Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F. & Strouboulis, J. (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. *European Molecular Biology Organisation Journal*, **24**, 2354-2366.
- Rooke, H.M. & Orkin, S.H. (2006) Phosphorylation of Gata1 at serine residues 72, 142, and 310 is not essential for hematopoiesis in vivo. *Blood*, **107**, 3527-3530.
- Roper, R.J. & Reeves, R.H. (2006) Understanding the basis for Down syndrome phenotypes. *PLoS Genetics*, **2**, e50.
- Rosenbauer, F., Koschmieder, S., Steidl, U. & Tenen, D.G. (2005) Effect of transcription-factor concentrations on leukemic stem cells. *Blood*, **106**, 1519-1524.
- Roth, D.B., Porter, T.N. & Wilson, J.H. (1985) Mechanisms of nonhomologous recombination in mammalian cells. *Molecular and Cellular Biology*, **5**, 2599-2607.



- Rutherford, T.R., Clegg, J.B. & Weatherall, D.J. (1979) K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin. *Nature*, **280**, 164-165.
- Rylski, M., Welch, J.J., Chen, Y.Y., Letting, D.L., Diehl, J.A., Chodosh, L.A., Blobel, G.A. & Weiss, M.J. (2003) GATA-1-mediated proliferation arrest during erythroid maturation. *Molecular and Cellular Biology*, **23**, 5031-5042.
- Sakamoto, H., Dai, G., Tsujino, K., Hashimoto, K., Huang, X., Fujimoto, T., Mucenski, M., Frampton, J. & Ogawa, M. (2006) Proper levels of c-Myb are discretely defined at distinct steps of hematopoietic cell development. *Blood*, **108**, 896-903.
- Salmon, P., Kindler, V., Ducrey, O., Chapuis, B., Zubler, R.H. & Trono, D. (2000) High-level transgene expression in human hematopoietic progenitors and differentiated blood lineages after transduction with improved lentiviral vectors. *Blood*, **96**, 3392-3398.
- Samokhvalov, I.M., Samokhvalova, N.I. & Nishikawa, S. (2007) Cell tracing shows the contribution of the yolk sac to adult haematopoiesis. *Nature*, **446**, 1056-1061.
- Sato, T., Toki, T., Kanezaki, R., Xu, G., Terui, K., Kanegane, H., Miura, M., Adachi, S., Migita, M., Morinaga, S., Nakano, T., Endo, M., Kojima, S., Kiyoi, H., Mano, H. & Ito, E. (2008) Functional analysis of JAK3 mutations in transient myeloproliferative disorder and acute megakaryoblastic leukaemia accompanying Down syndrome. *British Journal of Haematology*, **141**, 681-688.
- Saxena, R. & Kannan, M. (2008) Glanzmann's Thrombasthenia: An Overview. *Clinical and Applied Thrombosis Hemostasis*. (in press) epub Oct 16<sup>th</sup> 2008.
- Schafer, B.W., Blakely, B.T., Darlington, G.J. & Blau, H.M. (1990) Effect of cell history on response to helix  $\alpha$ -loop  $\alpha$ -helix family of myogenic regulators. *Nature*, **344**, 454-458.
- Scott, E.W., Simon, M.C., Anastasi, J. & Singh, H. (1994) Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science*, **265**, 1573-1577.
- Secco, P., Cotella, D. & Santoro, C. (2003) Selection of peptides with affinity for the N-terminal domain of GATA-1: Identification of a potential interacting protein. *Biochemical and Biophysical Research Communications*, **305**, 1061-1066.
- Shimada, A., Xu, G., Toki, T., Kimura, H., Hayashi, Y. & Ito, E. (2004) Fetal origin of the GATA1 mutation in identical twins with transient myeloproliferative disorder and acute megakaryoblastic leukemia accompanying Down syndrome. *Blood*, **103**, 366.
- Shimizu, R., Engel, J.D. & Yamamoto, M. (2008) GATA1-related leukaemias. *Nature Reviews Cancer*, **8**, 279-287.
- Shimizu, R., Kuroha, T., Ohneda, O., Pan, X., Ohneda, K., Takahashi, S., Philipsen, S. & Yamamoto, M. (2004) Leukemogenesis caused by incapacitated GATA-1 function. *Molecular and Cellular Biology*, **24**, 10814-10825.

- Shimizu, R., Takahashi, S., Ohneda, K., Engel, J.D. & Yamamoto, M. (2001) In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *European Molecular Biology Organisation Journal*, **20**, 5250-5260.
- Shimizu, R. & Yamamoto, M. (2005) Gene expression regulation and domain function of hematopoietic GATA factors. *Seminars in Cell and Developmental Biology*, **16**, 129-136.
- Shivdasani, R.A. (2001) Molecular and transcriptional regulation of megakaryocyte differentiation. *Stem Cells*, **19**, 397-407.
- Shivdasani, R.A., Fujiwara, Y., McDevitt, M.A. & Orkin, S.H. (1997) A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *European Molecular Biology Organisation Journal*, **16**, 3965-3973.
- Shivdasani, R.A. & Orkin, S.H. (1996) The transcriptional control of hematopoiesis [see comments]. *Blood*, **87**, 4025-4039.
- Siebert, P.D. & Fukuda, M. (1985) Regulation of glycophorin gene expression by a tumor-promoting phorbol ester in human leukemic K562 cells. *Journal of Biological Chemistry*, **260**, 640-645.
- Simon, M.C., Pevny, L., Wiles, M.V., Keller, G., Costantini, F. & Orkin, S.H. (1992) Rescue of erythroid development in gene targeted GATA-1- mouse embryonic stem cells. *Nature Genetics*, **1**, 92-98.
- Smith, D.L., Burthem, J. & Whetton, A.D. (2003) Molecular pathogenesis of chronic myeloid leukaemia. *Expert Reviews in Molecular Medicine*, **5**, 1-27.
- Socolovsky, M., Lodish, H.F. & Daley, G.Q. (1998) Control of hematopoietic differentiation: Lack of specificity in signaling by cytokine receptors. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6573-6575.
- Song, W.J., Sullivan, M.G., Legare, R.D., Hutchings, S., Tan, X., Kufrin, D., Ratajczak, J., Resende, I.C., Haworth, C., Hock, R., Loh, M., Felix, C., Roy, D.C., Busque, L., Kurnit, D., Willman, C., Gewirtz, A.M., Speck, N.A., Bushweller, J.H., Li, F.P., Gardiner, K., Poncz, M., Maris, J.M. & Gilliland, D.G. (1999) Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nature Genetics*, **23**, 166-175.
- Soriano, P. (1999) Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genetics*, **21**, 70-71.
- Stachura, D.L., Chou, S.T. & Weiss, M.J. (2006) Early block to erythromegakaryocytic development conferred by loss of transcription factor GATA-1. *Blood*, **107**, 87-97.
- Stamatoyannopoulos, G. (2005) Control of globin gene expression during development and erythroid differentiation. *Experimental Hematology*, **33**, 259-271.

- Stankiewicz, M.J. & Crispino, J.D. (2009) ETS2 and ERG promote megakaryopoiesis and synergize with alterations in GATA-1 to immortalize hematopoietic progenitor cells. *Blood (in press)* epub ahead of print Jan 23<sup>rd</sup> 2009.
- Stark, B., Jeison, M., Preudhomme, C., Fenaux, P., Ash, S., Korek, Y., Stein, J., Zaizov, R. & Yaniv, I. (2002) Acquired trisomy 21 and distinct clonal evolution in acute megakaryoblastic leukaemia in young monozygotic twins. *British Journal of Haematology*, **118**, 1082-1086.
- Stegmaier, P., Kel, A.E. & Wingender, E. (2004) Systematic DNA-binding domain classification of transcription factors. *Genome Informatics*, **15**, 276-286.
- Stopka, T., Amanatullah, D.F., Papetti, M. & Skoultchi, A.I. (2005) PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *European Molecular Biology Organisation Journal*, **24**, 3712-3723.
- Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. & Marra, M.A. (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences U S A*, **99**, 16899-16903.
- Suda, T., Suda, J. & Ogawa, M. (1984) Disparate differentiation in mouse hemopoietic colonies derived from paired progenitors. *Proceedings of the National Academy of Sciences U S A*, **81**, 2520-2524.
- Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. (2006) Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity*, **25**, 977-988.
- Suzuki, N., Suwabe, N., Ohneda, O., Obara, N., Imagawa, S., Pan, X., Motohashi, H. & Yamamoto, M. (2003) Identification and characterization of 2 types of erythroid progenitors that express GATA-1 at distinct levels. *Blood*, **102**, 3575-3583.
- Tabilio, A., Pelicci, P.G., Vinci, G., Mannoni, P., Civin, C.I., Vainchenker, W., Testa, U., Lipinski, M., Rochant, H. & Breton-Gorius, J. (1983) Myeloid and megakaryocytic properties of K-562 cell lines. *Cancer Research*, **43**, 4569-4574.
- Takahashi, S., Onodera, K., Motohashi, H., Suwabe, N., Hayashi, N., Yanai, N., Nabesima, Y. & Yamamoto, M. (1997) Arrest in primitive erythroid cell development caused by

promoter-specific disruption of the GATA-1 gene. *Journal of Biological Chemistry*, **272**, 12611-12615.

Taub, J.W., Huang, X., Matherly, L.H., Stout, M.L., Buck, S.A., Massey, G.V., Becton, D.L., Chang, M.N., Weinstein, H.J. & Ravindranath, Y. (1999) Expression of chromosome 21-localized genes in acute myeloid leukemia: differences between Down syndrome and non-Down syndrome blast cells and relationship to in vitro sensitivity to cytosine arabinoside and daunorubicin. *Blood*, **94**, 1393-1400.

Taub, J.W., Mundschau, G., Ge, Y., Poulik, J.M., Qureshi, F., Jensen, T., James, S.J., Matherly, L.H., Wechsler, J. & Crispino, J.D. (2004) Prenatal origin of GATA1 mutations may be an initiating step in the development of megakaryocytic leukemia in Down syndrome. *Blood*, **104**, 1588-1589.

Tavian, M., Hallais, M.F. & Peault, B. (1999) Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo. *Development*, **126**, 793-803.

Tavian, M., Robin, C., Coulombel, L. & Peault, B. (2001) The human embryo, but not its yolk sac, generates lympho-myeloid stem cells: mapping multipotent hematopoietic cell fate in intraembryonic mesoderm. *Immunity*, **15**, 487.

te Riele, H., Maandag, E.R. & Berns, A. (1992) Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences U S A*, **89**, 5128-5132.

Tetteroo, P.A., Massaro, F., Mulder, A., Schreuder-van Gelder, R. & von dem Borne, A.E. (1984) Megakaryoblastic differentiation of proerythroblastic K562 cell-line cells. *Leukemia Research*, **8**, 197-206.

Thomas, K.R. & Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, **51**, 503-512.

Till, J.E. & McCulloch, E. (1961) A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation Research*, **14**, 213-222.

Tober, J., Koniski, A., McGrath, K.E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K.K., Waugh, R. & Palis, J. (2007) The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood*, **109**, 1433-1441.

Toki, T., Katsuoka, F., Kanezaki, R., Xu, G., Kurotaki, H., Sun, J., Kamio, T., Watanabe, S., Tandai, S., Terui, K., Yagihashi, S., Komatsu, N., Igarashi, K., Yamamoto, M. & Ito, E. (2005) Transgenic expression of BACH1 transcription factor results in megakaryocytic impairment. *Blood*, **105**, 3100-3108.

Towatari, M., Ciro, M., Ottolenghi, S., Tsuzuki, S. & Enver, T. (2004) Involvement of mitogen-activated protein kinase in the cytokine-regulated phosphorylation of transcription factor GATA-1. *Hematology Journal*, **5**, 262-272.

- Trimborn, T., Gribnau, J., Grosveld, F. & Fraser, P. (1999) Mechanisms of developmental control of transcription in the murine alpha - and beta -globin loci. *Genes and Development*, **13**, 112-124.
- Tronik-Le Roux, D., Roullot, V., Poujol, C., Kortulewski, T., Nurden, P. & Marguerie, G. (2000) Thrombasthenic mice generated by replacement of the integrin alpha(IIb) gene: demonstration that transcriptional activation of this megakaryocytic locus precedes lineage commitment. *Blood*, **96**, 1399-1408.
- Tsai, S.F., Martin, D.I., Zon, L.I., D'Andrea, A.D., Wong, G.G. & Orkin, S.H. (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, **339**, 446-451.
- Tsai, S.F., Strauss, E. & Orkin, S.H. (1991) Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes and Development*, **5**, 919-931.
- Tsang, A.P., Fujiwara, Y., Hom, D.B. & Orkin, S.H. (1998) Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes and Development*, **12**, 1176-1188.
- Tsang, A.P., Visvader, J.E., Turner, C.A., Fujiwara, Y., Yu, C., Weiss, M.J., Crossley, M. & Orkin, S.H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*, **90**, 109-119.
- Tsuzuki, S., Hong, D., Gupta, R., Matsuo, K., Seto, M. & Enver, T. (2007) Isoform-specific potentiation of stem and progenitor cell engraftment by AML1/RUNX1. *PLoS Medicine*, **4**, e172.
- Tunstall-Pedoe, O., Roy, A., Karadimitris, A., de la Fuente, J., Fisk, N.M., Bennett, P., Norton, A., Vyas, P. & Roberts, I. (2008) Abnormalities in the myeloid progenitor compartment in Down syndrome fetal liver precede acquisition of GATA1 mutations. *Blood*, **112**, 4507-11.
- Ueki, N., Zhang, L. & Hayman, M.J. (2004) Ski negatively regulates erythroid differentiation through its interaction with GATA1. *Molecular and Cellular Biology*, **24**, 10118-10125.
- Vainchenker, W., Testa, U., Guichard, J., Titeux, M. & Breton-Gorius, J. (1981) Heterogeneity in the cellular commitment of a human leukemic cell line: K 562. *Blood Cells*, **7**, 357-375.
- Valverde-Garduno, V., Guyot, B., Anguita, E., Hamlett, I., Porcher, C. & Vyas, P. (2004) Differences in the chromatin structure and cis-element organization of the human and mouse GATA1 loci: implications for cis-element identification. *Blood*, **104**, 3106-3116.
- Visvader, J.E., Crossley, M., Hill, J., Orkin, S.H. & Adams, J.M. (1995) The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Molecular and Cellular Biology*, **15**, 634-641.

- Visvader, J.E., Elefanty, A.G., Strasser, A. & Adams, J.M. (1992) GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *European Molecular Biology Organisation Journal*, **11**, 4557-4564.
- Vyas, P., Ault, K., Jackson, C.W., Orkin, S.H. & Shivdasani, R.A. (1999a) Consequences of GATA-1 deficiency in megakaryocytes and platelets. *Blood*, **93**, 2867-2875.
- Vyas, P., McDevitt, M.A., Cantor, A.B., Katz, S.G., Fujiwara, Y. & Orkin, S.H. (1999b) Different sequence requirements for expression in erythroid and megakaryocytic cells within a regulatory element upstream of the GATA-1 gene. *Development*, **126**, 2799-2811.
- Vyas, P. & Roberts, I. (2006) Down myeloid disorders: a paradigm for childhood preleukaemia and leukaemia and insights into normal megakaryopoiesis. *Early Human Development*, **82**, 767-773.
- Wall, L., deBoer, E. & Grosveld, F. (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes and Development*, **2**, 1089-1100.
- Walters, D.K., Mercher, T., Gu, T.L., O'Hare, T., Tyner, J.W., Loriaux, M., Goss, V.L., Lee, K.A., Eide, C.A., Wong, M.J., Stoffregen, E.P., McGreevey, L., Nardone, J., Moore, S.A., Crispino, J., Boggon, T.J., Heinrich, M.C., Deininger, M.W., Polakiewicz, R.D., Gilliland, D.G. & Druker, B.J. (2006) Activating alleles of JAK3 in acute megakaryoblastic leukemia. *Cancer Cell*, **10**, 65-75.
- Waltzer, L., Ferjoux, G., Bataille, L. & Haenlin, M. (2003) Cooperation between the GATA and RUNX factors Serpent and Lozenge during *Drosophila* hematopoiesis. *European Molecular Biology Organisation Journal*, **22**, 6516-6525.
- Wechsler, J., Greene, M., McDevitt, M.A., Anastasi, J., Karp, J.E., Le Beau, M.M. & Crispino, J.D. (2002) Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. *Nature Genetics*, **32**, 148-152.
- Weiss, M.J., Keller, G. & Orkin, S.H. (1994) Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes and Development*, **8**, 1184-1197.
- Weiss, M.J. & Orkin, S.H. (1995) Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proceedings of the National Academy of Sciences U S A*, **92**, 9623-9627.
- Weiss, M.J., Yu, C. & Orkin, S.H. (1997) Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Molecular and Cellular Biology*, **17**, 1642-1651.
- Welch, J.J., Watts, J.A., Vakoc, C.R., Yao, Y., Wang, H., Hardison, R.C., Blobel, G.A., Chodosh, L.A. & Weiss, M.J. (2004) Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*, **104**, 3136-3147.
- Whyatt, D., Lindeboom, F., Karis, A., Ferreira, R., Milot, E., Hendriks, R., de Bruijn, M., Langeveld, A., Gribnau, J., Grosveld, F. & Philipsen, S. (2000) An intrinsic but cell-

nonautonomous defect in GATA-1-overexpressing mouse erythroid cells. *Nature*, **406**, 519-524.

Wiemels, J.L., Xiao, Z., Buffler, P.A., Maia, A.T., Ma, X., Dicks, B.M., Smith, M.T., Zhang, L., Feusner, J., Wiencke, J., Pritchard-Jones, K., Kempinski, H. & Greaves, M. (2002) In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. *Blood*, **99**, 3801-3805.

Woods, N.B., Mikkola, H., Nilsson, E., Olsson, K., Trono, D., Karlsson, S. (2001) Lentiviral-mediated gene transfer into haematopoietic stem cells. *Journal of Internal Medicine*, **249**, 339-343.

Wu, H. & Sun, Y.E. (2006) Epigenetic regulation of stem cell differentiation. *Pediatric Research*, **59**, 21R-25R.

Wu, S., Ying, G., Wu, Q. & Capecchi, M.R. (2008) A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. *Nature Protocols*, **3**, 1056-1076.

Xiao, W. & Oefner, P.J. (2001) Denaturing high-performance liquid chromatography: A review. *Human Mutation*, **17**, 439-474.

Xu, G., Kato, K., Toki, T., Takahashi, Y., Terui, K. & Ito, E. (2006) Development of acute megakaryoblastic leukemia from a minor clone in a Down syndrome patient with clinically overt transient myeloproliferative disorder. *Journal of Pediatric Hematology Oncology*, **28**, 696-698.

Xu, G., Nagano, M., Kanezaki, R., Toki, T., Hayashi, Y., Taketani, T., Taki, T., Mitui, T., Koike, K., Kato, K., Imaizumi, M., Sekine, I., Ikeda, Y., Hanada, R., Sako, M., Kudo, K., Kojima, S., Ohneda, O., Yamamoto, M. & Ito, E. (2003) Frequent mutations in the GATA-1 gene in the transient myeloproliferative disorder of Down syndrome. *Blood*, **102**, 2960-2968.

Xu, M.J., Matsuoka, S., Yang, F.C., Ebihara, Y. & Manabe, A. (2001) Evidence for the presence of murine primitive megakaryocytopoiesis in the early yolk sac. *Blood*, **97**, 2016.

Yamada, O., Takanashi, M., Ujihara, M. & Mizoguchi, H. (1998) Down-regulation of telomerase activity is an early event of cellular differentiation without apparent telomeric DNA change. *Leukemia Research*, **22**, 711-717.

Yoder, M.C., Hiatt, K., Dutt, P., Mukherjee, P., Bodine, D.M. & Orlic, D. (1997) Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity*, **7**, 335.

Yokota, T., Huang, J., Tavian, M., Nagai, Y. & Hirose, J. (2006) Tracing the first waves of lymphopoiesis in mice. *Development*, **133**, 2041.

Yoshida, N., Aizu-Yokota, E., Sonoda, Y., Moriwaki, Y., Kishi, K. & Kasahara, T. (2007) Production and regulation of eotaxin-2/CCL24 in a differentiated human leukemic cell line, HT93. *Biological and Pharmaceutical Bulletin*, **30**, 1826-1832.

- Yu, C., Cantor, A.B., Yang, H., Browne, C., Wells, R.A., Fujiwara, Y. & Orkin, S.H. (2002a) Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo. *Journal of Experimental Medicine*, **195**, 1387-1395.
- Yu, C., Niakan, K.K., Matsushita, M., Stamatoyannopoulos, G., Orkin, S.H. & Raskind, W.H. (2002b) X-linked thrombocytopenia with thalassemia from a mutation in the amino finger of GATA-1 affecting DNA binding rather than FOG-1 interaction. *Blood*, **100**, 2040-2045.
- Yu, Y.L., Chiang, Y.J., Chen, Y.C., Papetti, M., Juo, C.G., Skoultschi, A.I. & Yen, J.J. (2005) MAPK-mediated phosphorylation of GATA-1 promotes Bcl-XL expression and cell survival. *Journal of Biological Chemistry*, **280**, 29533-29542.
- Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G. & Soriano, P. (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proceedings of the National Academy of Sciences U S A*, **94**, 3789-3794.
- Zeigler, B.M., Sujiyama, D., Chen, M., Guo, Y., Downd, K.M. & Speck, N.A. (2006) The allantois and chorion, when isolated before circulation or chorio-allantoic fusion, have hematopoietic potential. *Development*, **133**, 4183.
- Zhan, M., Miller, C.P., Papayannopoulou, T., Stamatoyannopoulos, G. & Song, C.Z. (2007) MicroRNA expression dynamics during murine and human erythroid differentiation. *Experimental Hematology*, **35**, 1015-1025.
- Zhang, S. (2008) The role of aberrant transcription factor in the progression of chronic myeloid leukemia. *Leukemia and Lymphoma*, **49**, 1463-1469.
- Zheng, J., Kitajima, K., Sakai, E., Kimura, T., Minegishi, N., Yamamoto, M. & Nakano, T. (2006) Differential effects of GATA-1 on proliferation and differentiation of erythroid lineage cells. *Blood*, **107**, 520-527.
- Zipursky, A. (2003) Transient leukaemia--a benign form of leukaemia in newborn infants with trisomy 21. *British Journal of Haematology*, **120**, 930-938.
- Zipursky, A., Brown, E., Christensen, H., Sutherland, R. & Doyle, J. (1997) Leukemia and/or myeloproliferative syndrome in neonates with Down syndrome. *Seminars in Perinatology*, **21**, 97-101.
- Zon, L.I., Tsai, S.F., Burgess, S., Matsudaira, P., Bruns, G.A. & Orkin, S.H. (1990) The major human erythroid DNA-binding protein (GF-1): primary sequence and localization of the gene to the X chromosome. *Proceedings of the National Academy of Sciences U S A*, **87**, 668-672.



